

THESIS / THÈSE

MASTER EN SCIENCES MATHÉMATIQUES

Trois algorithmes de recherche d'un zéro d'un opérateur maximal monotone, basés sur la notion d'e-élargissement

Weis, Alain

Award date:
1999

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



Facultés Universitaires Notre-Dame de la Paix
Namur
Faculté des Sciences - Département de Mathématique

1/40
41
45
5/51
5/54
5/58

Trois algorithmes de recherche d'un zéro d'un opérateur maximal monotone, basés sur la notion d' ε -élargissement

Mémoire présenté pour l'obtention du grade
de Licencié en Sciences
mathématiques
par

Promoteur : Jean-Jacques STRODIOT

Alain WEIS

Année académique : 1998-1999

Remerciements

En premier lieu, je voudrais remercier le Professeur J.-J. STRODIOT d'avoir bien voulu être le promoteur de ce mémoire. La qualité de la documentation qu'il m'a fournie, sa disponibilité tout au long de cette année académique 1998-1999 et ses remarques constructives ont permis de mener à bien ce travail.

Je remercie également les Professeurs F.M. CALLIER et V.H. NGUYEN d'avoir accepté la tâche de lecteur.

Ma gratitude s'adresse aussi à Madame A. LEMAÎTRE pour les nombreux services qu'elle m'a rendus durant mes années d'études à Namur.

Finalement, je tiens à remercier Madame M. VAN CAENEGEM pour le soin et la rapidité qu'elle a apportés à la dactylographie de ce mémoire.

Alain WEIS.

Résumé

Étant donné un opérateur maximal monotone T , on définit un ε -élargissement T^ε dont on établit les principales propriétés théoriques, notamment une formule de transport, la Lipschitz continuité et une généralisation du Théorème de Brønsted et Rockafellar. Cet ε -élargissement permet de décrire un premier algorithme pour la recherche d'un zéro de T . En se basant sur la méthode faisceau de l'optimisation, on remplace ensuite les ensembles convexes $T^\varepsilon(x)$ par une approximation polyédrale permettant de donner une version implémentable de l'algorithme. On montre que les deux algorithmes sont convergents. Finalement, on présente une méthode hybride effectuant à chaque itération un pas de type extragradient approché et ceci en utilisant des informations obtenues par une solution approchée d'un sous-problème prox qui fait intervenir à la fois un terme reste et l' ε -élargissement T^ε . Sous certaines hypothèses, on établit la convergence faible globale et une vitesse de convergence linéaire locale.

Summary

Given a maximal monotone operator T , we define an ε -enlargement T^ε of which we establish some theoretical properties, including a transportation formula, Lipschitz continuity, and a result generalizing Brønsted and Rockefellar's theorem. We use this ε -enlargement to define a first algorithm for finding a zero of T . Then we replace the convex sets $T^\varepsilon(x)$ by a polyhedral approximation following the bundle strategy of nonsmooth optimization, which allows us to develop an implementable version of the algorithm. The two algorithms are proved to be convergent. Finally we present a hybrid algorithm performing at each iteration an approximate extragradient-type step using information obtained from an approximate solution of a proximal point subproblem which is handled both through an error term and the ε -enlargement T^ε . Under suitable assumptions we establish weak global convergence and local linear rate of convergence.

Introduction

Ce mémoire a pour objet d'étudier le problème de la recherche d'un zéro d'un opérateur maximal monotone. Il s'agit d'un problème très abstrait qui présente l'avantage de couvrir un nombre assez large de problèmes plus concrets, qui peuvent alors être vus comme des cas particuliers du premier. Ainsi, le problème de recherche d'un minimum d'une fonction convexe non différentiable, des problèmes de min-max, des problèmes de complémentarité et des problèmes aux inéquations variationnelles (VIP) tombent dans ce cadre général.

Un premier chapitre est consacré à l'introduction des concepts d'opérateur maximal monotone et d' ε -élargissement, ainsi qu'à la présentation de certaines propriétés de ces notions. Comme ce sont des notions très générales, on essaie tout au long du mémoire de garder le parallélisme avec le cas particulier du sous-différentiel d'une fonction convexe, car ce domaine plus largement connu permettra sans doute de suivre plus facilement les raisonnements faits dans le cadre général. Partant de ces propriétés, on présente au chapitre suivant en détail un article de Burachik, Sagastizábal et Svaiter [3], pour trouver le zéro d'un opérateur maximal monotone, basé sur des projections sur des sous-espaces bien choisis. Cet algorithme étend en fait les méthodes d' ε -descente de l'optimisation non-différentiable, en travaillant avec le vecteur de norme minimum dans $T^{\varepsilon_k}(x^k)$ (avec $\varepsilon_k \rightarrow 0$).

Au Chapitre 3, on quitte momentanément le cadre purement abstrait pour présenter une version implémentable de l'algorithme, conçue par les mêmes personnes [4], qui approxime les ensembles convexes $T^\varepsilon(x)$ en généralisant la méthode faisceau de l'optimisation.

Le quatrième chapitre, basé sur un article de Solodov et Svaiter [15], fait appel aux méthodes "prox" pour trouver le zéro d'un opérateur maximal monotone. Ce qui est innovateur, c'est que, premièrement, on introduit à la fois un terme reste et on travaille avec un ε -élargissement de l'opérateur; et deuxièmement, on remplace, lors de l'évaluation approximative des itérés, la condition classique de sommabilité a priori, par une condition sur le reste qui peut être évaluée à chaque itération.

Même si, pour cet algorithme, on n'a pas été jusqu'à présenter une version implémentable, son cadre général permet quand même la conception de nouveaux algorithmes, notamment une méthode de type Newton globalement convergente.

Chapitre 1

Opérateurs maximaux monotones

1.1 Définitions et propriétés

Définition 1.1

Soit X un espace de Hilbert sur \mathbb{R} .

Un **opérateur** T est une application de X dans $\mathcal{P}(X)$, ensemble des parties de X , notée $T : X \rightarrow \mathcal{P}(X)$.

On identifie T avec son graphe

$$G(T) = \{(x, y) \mid y \in T(x)\} \subseteq X \times X.$$

Il est important de noter qu'on a écrit $y \in T(x)$ et non pas $y = T(x)$ car, à un point $x \in X$, un opérateur peut associer un ensemble de points de X . On dit que l'opérateur est **multivoque**. Si, par contre, pour chaque point $x \in X$, $T(x)$ contient au plus un élément, on parle d'opérateur **univoque** (ou fonctionnel).

Définition 1.2

Le **domaine** d'un opérateur $T : X \rightarrow \mathcal{P}(X)$ est l'ensemble

$$\begin{aligned} D(T) &\stackrel{\text{déf}}{=} \{x \in X \mid T(x) \neq \emptyset\} \\ &= \{x \in X \mid \exists y \in X \quad (x, y) \in G(T)\}. \end{aligned}$$

Définition 1.3

L'image d'un opérateur $T : X \rightarrow \mathcal{P}(X)$ est l'ensemble

$$\begin{aligned} R(T) &\stackrel{\text{déf}}{=} \{y \in X \mid \exists x \in X \quad (x, y) \in G(T)\} \\ &= \bigcup_{x \in X} T(x) \\ &\stackrel{\text{not}}{=} T(X) . \end{aligned}$$

Définition 1.4

L'inverse d'un opérateur $T : X \rightarrow \mathcal{P}(X)$ est l'opérateur $T^{-1} : X \rightarrow \mathcal{P}(X)$ dont le graphe est symétrique de celui de T , c'est-à-dire

$$(y, x) \in G(T^{-1}) \Leftrightarrow (x, y) \in G(T)$$

ou encore

$$x \in T^{-1}(y) \Leftrightarrow y \in T(x) .$$

Contrairement à l'inverse d'une fonction, l'inverse d'un opérateur existe toujours. On a immédiatement

$$(T^{-1})^{-1} = T$$

ainsi que

$$D(T^{-1}) = R(T) .$$

Définition 1.5

Soit X un espace de Hilbert muni d'un produit scalaire noté $\langle \cdot, \cdot \rangle$ dont on peut déduire une norme notée $\| \cdot \|$.

On dit que l'opérateur $T : X \rightarrow \mathcal{P}(X)$ est **monotone** si et seulement si :

$$\left. \begin{array}{l} \forall x_1, x_2 \in D(T) , \\ \forall y_1 \in T(x_1) , \forall y_2 \in T(x_2) , \end{array} \right\} \quad \langle x_1 - x_2, y_1 - y_2 \rangle \geq 0 .$$

Si l'opérateur est univoque, on écrit

$$\forall x_1, x_2 \in D(T) : \quad \langle x_1 - x_2, T(x_1) - T(x_2) \rangle \geq 0 .$$

Remarque : il s'agit en fait de la généralisation de la notion de fonction croissante. En effet, $f : \mathbb{R} \rightarrow \mathbb{R}$ est croissante si et seulement si

$$\forall x_1, x_2 : x_1 \leq x_2 \Rightarrow f(x_1) \leq f(x_2)$$

ou encore :

$$\forall x_1, x_2 : (x_1 - x_2) (f(x_1) - f(x_2)) \geq 0 .$$

Dû à la symétrie du produit scalaire, si T est monotone, T^{-1} l'est aussi.

Proposition 1.1

Soit $T : X \rightarrow \mathcal{P}(X)$.

T est monotone si et seulement si

$$\left. \begin{array}{l} \forall x_1, x_2 \in D(T) , \\ \forall y_1 \in T(x_1) , \forall y_2 \in T(x_2) , \\ \forall \lambda > 0 , \end{array} \right\} \quad \|x_1 - x_2\| \leq \|(x_1 - x_2) + \lambda (y_1 - y_2)\| .$$

Preuve :

– La condition est nécessaire : en effet, T étant monotone et $\lambda > 0$, on a

$$\begin{aligned} & \| (x_1 - x_2) + \lambda (y_1 - y_2) \|^2 \\ &= \|x_1 - x_2\|^2 + \lambda^2 \|y_1 - y_2\|^2 + 2\lambda \langle x_1 - x_2, y_1 - y_2 \rangle \\ &\geq \|x_1 - x_2\|^2 , \end{aligned}$$

et donc

$$\| (x_1 - x_2) + \lambda (y_1 - y_2) \| \geq \|x_1 - x_2\| .$$

– La condition est aussi suffisante : en effet, si on a

$$\|x_1 - x_2\| \leq \| (x_1 - x_2) + \lambda (y_1 - y_2) \| ,$$

en élevant au carré et en développant comme dans la première partie, on obtient

$$\lambda^2 \|y_1 - y_2\|^2 + 2 \lambda \langle x_1 - x_2, y_1 - y_2 \rangle \geq 0$$

et en divisant par $\lambda > 0$:

$$\lambda \|y_1 - y_2\|^2 + 2 \langle x_1 - x_2, y_1 - y_2 \rangle \geq 0 .$$

Il suffit alors de faire tendre λ vers 0 , et de diviser par 2 pour obtenir

$$\langle x_1 - x_2, y_1 - y_2 \rangle \geq 0 ,$$

c'est-à-dire T est monotone. ■

Cette propriété signifie en fait que

$$\left. \begin{array}{l} \forall x_1, x_2 \in D(T) , \\ \forall u_1 \in (I + \lambda T)(x_1) , \forall u_2 \in (I + \lambda T)(x_2) , \\ \forall \lambda > 0 , \end{array} \right\} \quad \|x_1 - x_2\| \leq \|u_1 - u_2\| ,$$

c'est-à-dire que l'opérateur

$$(I + \lambda T)^{-1} \stackrel{\text{not}}{=} J_\lambda$$

appelé **résolvante de T** , est, pour tout $\lambda > 0$, une contraction de $R(I + \lambda T) \rightarrow X$,
ou autrement dit, pour tout $y \in X$, l'équation $x + \lambda T(x) \ni y$ admet **au plus** une solution.

On va maintenant s'intéresser aux opérateurs pour lesquels cette équation possède **exactement** une solution.

Définition 1.6

On dit que l'opérateur $T : X \rightarrow \mathcal{P}(X)$ est **maximal monotone** si et seulement si T est monotone et, pour tout opérateur monotone S tel que

$$T(x) \subset S(x) , \quad \forall x \quad \Rightarrow \quad T = S .$$

Autrement dit, son graphe est maximal pour l'inclusion dans l'ensemble des graphes monotones ou encore :

T est **maximal monotone** si et seulement si T est monotone et,
pour tout $(x, y) \in X \times X$ tel que $\langle y - \eta, x - \xi \rangle \geq 0$, $\forall \xi \in D(T)$, $\forall \eta \in T(\xi)$,
alors $y \in T(x)$.

Dû à la symétrie du produit scalaire, T est maximal monotone si et seulement si T^{-1} est maximal monotone.

Pour les opérateurs maximaux monotones, on a une caractérisation très importante :

Théorème 1.1

Soit $T : X \rightarrow \mathcal{P}(X)$ un opérateur.

Alors les trois assertions suivantes sont équivalentes :

1. *T est maximal monotone.*
2. *T est monotone et $R(I + T) = X$.*
3. *$\forall \lambda > 0 : J_\lambda \equiv (I + \lambda T)^{-1}$ est une contraction définie sur X entier.*

Preuve : voir [1]. ■

Par la définition, il est évident que λT (avec $\lambda > 0$) est maximal monotone si T l'est aussi. Par le point 2. du théorème, l'équation $x + \lambda T(x) \ni y$ admet donc **au moins** une solution. Comme on sait déjà qu'elle en admet **au plus** une, elle en admet **une et une seule** comme mentionné déjà plus haut.

Ceci permet d'écrire le point 3. du théorème sous la forme :

*La résolvante $J_\lambda \equiv (I + \lambda T)^{-1}$ d'un opérateur maximal monotone T sur X est **univoque** contractante et partout définie.*

Donnons maintenant deux définitions qui nous permettront de présenter un exemple d'opérateur multivoque maximal monotone.

Définition 1.7

Soit $f : \mathbb{R}^n \rightarrow]-\infty, +\infty]$ une fonction convexe non nécessairement différentiable.

Un vecteur $x^* \in \mathbb{R}^n$ est un **sous-gradient** de f au point $x \in \mathbb{R}^n$ si et seulement si

$$f(z) \geq f(x) + \langle x^*, z - x \rangle, \quad \forall z \in \mathbb{R}^n .$$

Dans le cas différentiable, le gradient est un sous-gradient. Un sous-gradient x^* de f en x est une généralisation du gradient dans le sens qu'il détermine une fonction affine dont le graphe est un hyperplan qui se situe en dessous de l'épigraphe de f en passant par $(x, f(x))$.

Définition 1.8

On appelle **sous-différentiel** de f en x l'ensemble de tous les sous-gradients de f en x :

$$\partial f(x) \stackrel{\text{d\'ef}}{=} \{x^* \in \mathbb{R}^n \mid f(z) \geq f(x) + \langle x^*, z - x \rangle, \quad \forall z \in \mathbb{R}^n\}.$$

Considérons maintenant l'opérateur (éventuellement multivoque)

$$\begin{aligned} \partial f : \quad \mathbb{R}^n &\rightarrow \mathcal{P}(\mathbb{R}^n), \\ x &\mapsto \partial f(x), \end{aligned}$$

où $f : \mathbb{R}^n \rightarrow]-\infty, +\infty]$ est une fonction convexe propre.

Alors l'opérateur différentiel ∂f est monotone. En effet : soient $x_1, x_2 \in \mathbb{R}^n$.

Par définition du sous-différentiel, on a

$$x_1^* \in \partial f(x_1) \quad \Leftrightarrow \quad f(z) \geq f(x_1) + \langle x_1^*, z - x_1 \rangle, \quad \forall z, \quad (1.1)$$

$$x_2^* \in \partial f(x_2) \quad \Leftrightarrow \quad f(z) \geq f(x_2) + \langle x_2^*, z - x_2 \rangle, \quad \forall z. \quad (1.2)$$

Donc en particulier pour $z = x_2$ dans (1.1) et $z = x_1$ dans (1.2), on obtient

$$\begin{aligned} f(x_2) &\geq f(x_1) + \langle x_1^*, x_2 - x_1 \rangle, \\ f(x_1) &\geq f(x_2) + \langle x_2^*, x_1 - x_2 \rangle. \end{aligned}$$

En sommant ces deux inégalités, on a :

$$f(x_2) + f(x_1) \geq f(x_1) + f(x_2) + \langle x_1^*, x_2 - x_1 \rangle + \langle x_2^*, x_1 - x_2 \rangle$$

c'est-à-dire

$$0 \geq -\langle x_1^*, x_1 - x_2 \rangle + \langle x_2^*, x_1 - x_2 \rangle$$

ou encore

$$0 \leq \langle x_1^* - x_2^*, x_1 - x_2 \rangle.$$

Ceci montre que l'opérateur ∂f est monotone. ■

Proposition 1.2

Si la fonction f est convexe, propre et semi-continue inférieurement, alors l'opérateur ∂f est maximal monotone.

Pour démontrer cette proposition, on a besoin d'un lemme préliminaire.

Lemme 1.1

Soit ϕ une fonction convexe propre sur X et $\alpha \geq 0$.

La fonction convexe

$$\psi : x \mapsto \phi(x) + \frac{1}{2} \alpha \|x - y\|^2$$

atteint son minimum en x_0 si et seulement si $\alpha(y - x_0) \in \partial\phi(x_0)$.

Preuve :

– Montrons d'abord que la condition est suffisante :

si $\alpha(y - x_0) \in \partial\phi(x_0)$, alors $\phi(x_0) < +\infty$ et on a, par définition du sous-différentiel

$$\forall \xi \in X : \quad \phi(\xi) - \phi(x_0) \geq \alpha \langle y - x_0, \xi - x_0 \rangle .$$

Or, on a l'inégalité suivante

$$\alpha \langle y - x_0, \xi - x_0 \rangle \geq \frac{\alpha}{2} [\|x_0 - y\|^2 - \|\xi - y\|^2]$$

car

$$\begin{aligned} \langle y - x_0, \xi - x_0 \rangle &\geq \frac{1}{2} [\langle x_0 - y, x_0 - y \rangle - \langle \xi - y, \xi - y \rangle] \\ \Leftrightarrow \quad \langle y, \xi \rangle - \langle y, x_0 \rangle - \langle x_0, \xi \rangle + \|x_0\|^2 \\ &\geq \frac{1}{2} [\|x_0\|^2 - 2\langle x_0, y \rangle + \|y\|^2 - \|\xi\|^2 + 2\langle \xi, y \rangle - \|y\|^2] \\ \Leftrightarrow \quad \langle y, \xi \rangle - \langle y, x_0 \rangle - \langle x_0, \xi \rangle + \|x_0\|^2 \\ &\geq \frac{1}{2} \|x_0\|^2 - \langle y, x_0 \rangle - \frac{1}{2} \|\xi\|^2 + \langle y, \xi \rangle \\ \Leftrightarrow \quad -\langle x_0, \xi \rangle + \frac{1}{2} \|x_0\|^2 + \frac{1}{2} \|\xi\|^2 &\geq 0 \\ \Leftrightarrow \quad \|x_0 - \xi\|^2 &\geq 0 \end{aligned} \quad \text{ce qui est toujours vrai.}$$

On obtient alors les relations équivalentes suivantes

$$\begin{aligned} \forall \xi \in X : \quad \phi(\xi) - \phi(x_0) &\geq \frac{\alpha}{2} [\|x_0 - y\|^2 - \|\xi - y\|^2] , \\ \forall \xi \in X : \quad \phi(\xi) + \frac{\alpha}{2} \|\xi - y\|^2 - \phi(x_0) - \frac{\alpha}{2} \|x_0 - y\|^2 &\geq 0 , \\ \forall \xi \in X : \quad \psi(\xi) - \psi(x_0) &\geq 0 , \end{aligned}$$

qui expriment que la fonction ψ atteint son minimum en x_0 .

– Montrons ensuite que la condition est également nécessaire.

Soit $\eta \in X$ et posons

$$\xi = (1 - t) x_0 + t \eta, \quad t \in]0, 1[.$$

Comme ϕ est convexe, on a successivement

$$\phi(\xi) \leq t \phi(\eta) + (1 - t) \phi(x_0)$$

$$\phi(\xi) - \phi(x_0) \leq t \phi(\eta) - t \phi(x_0)$$

$$t [\phi(\eta) - \phi(x_0)] \geq \phi(\xi) - \phi(x_0).$$

Or,

$$\phi(\xi) - \phi(x_0) \geq \frac{\alpha}{2} \|x_0 - y\|^2 - \frac{\alpha}{2} \|(1 - t) x_0 + t \eta - y\|^2,$$

car, si ce n'était pas le cas, on aurait

$$\phi(\xi) - \phi(x_0) < \frac{\alpha}{2} \|x_0 - y\|^2 - \frac{\alpha}{2} \|(1 - t) x_0 + t \eta - y\|^2$$

$$\Leftrightarrow \phi(\xi) + \frac{\alpha}{2} \|\xi - y\|^2 < \phi(x_0) + \frac{\alpha}{2} \|x_0 - y\|^2$$

$$\Leftrightarrow \psi(\xi) < \psi(x_0)$$

ce qui est impossible car, par hypothèse, x_0 est le minimum de ψ .

On peut donc continuer la minoration :

$$\begin{aligned} t [\phi(\eta) - \phi(x_0)] &\geq \frac{\alpha}{2} \|x_0 - y\|^2 - \frac{\alpha}{2} \|(1 - t) x_0 + t \eta - y\|^2 \\ &= \frac{\alpha}{2} \|x_0 - y\|^2 - \frac{\alpha}{2} \|x_0 - y\|^2 \\ &\quad - 2 \frac{\alpha}{2} \langle x_0 - y, t (\eta - x_0) \rangle - \frac{\alpha}{2} \|t (\eta - x_0)\|^2 \\ &= -\alpha t \langle x_0 - y, \eta - x_0 \rangle - \frac{\alpha}{2} t^2 \|\eta - x_0\|^2. \end{aligned}$$

Divisant par $t > 0$ et faisant tendre t vers 0, on obtient

$$\phi(\eta) - \phi(x_0) \geq \alpha \langle y - x_0, \eta - x_0 \rangle$$

ou encore

$$\phi(\eta) \geq \phi(x_0) + \langle \alpha (y - x_0), \eta - x_0 \rangle.$$

Comme $\eta \in X$ est arbitraire, on en déduit $\alpha (y - x_0) \in \partial\phi(x_0)$. ■

On peut maintenant prouver que ∂f est maximal monotone si f est convexe, propre et semi-continue inférieurement.

Preuve : Par la caractérisation du Théorème 1.1, il suffit de montrer que ∂f est monotone et $R(I + \partial f) = X$.

Comme on sait déjà que ∂f est monotone, montrons que $R(I + \partial f) = X$, c'est-à-dire

$$\forall y \in X : \exists x_0 \text{ tel que } y \in x_0 + \partial f(x_0).$$

Grâce au Lemme 1.1 (avec $\alpha = 1$), il suffit de montrer que la fonction

$$\psi(x) : x \mapsto f(x) + \frac{1}{2} \|x - y\|^2$$

atteint son minimum en un point x_0 de X .

La fonction f , étant convexe, est minorée par une fonction affine (en vertu du Théorème de Hahn-Banach). Dès lors, $\psi(x) \rightarrow +\infty$ lorsque $\|x\| \rightarrow +\infty$, c'est-à-dire ψ est coercive sur X . Comme ψ est aussi semi-continue inférieurement, elle atteint son minimum en un point x_0 de X . ■

Cet exemple d'opérateur maximal monotone ∂f est très important car, dans tout ce qui suit, on va essayer de généraliser aux opérateurs maximaux monotones ce qui a été fait pour le sous-différentiel en optimisation.

Avant d'aborder cette généralisation, rappelons la définition suivante.

Définition 1.9

Soit $f : \mathbb{R}^n \rightarrow]-\infty, +\infty]$ une fonction convexe propre. Soient $x \in \mathbb{R}^n$ et $\varepsilon \geq 0$.

On appelle ε -sous-différentiel de f en x l'ensemble

$$\partial_\varepsilon f(x) \stackrel{\text{déf}}{=} \{x^* \in \mathbb{R}^n \mid f(z) \geq f(x) - \varepsilon + \langle x^*, z - x \rangle, \quad \forall z \in \mathbb{R}^n\}.$$

Comme $\varepsilon \geq 0$, il suit immédiatement de la définition que

$$\begin{aligned} \partial f(x) &= \partial_0 f(x) \subseteq \partial_\varepsilon f(x) \\ \text{et} \quad \varepsilon_1 \leq \varepsilon_2 &\Rightarrow \partial_{\varepsilon_1} f(x) \subseteq \partial_{\varepsilon_2} f(x) \\ \text{ainsi que} \quad \bigcap_{\varepsilon > 0} \partial_\varepsilon f(x) &= \partial f(x). \end{aligned}$$

Cet élargissement du sous-différentiel permet de trouver une approximation de $\partial f(x)$ dont le calcul exact n'est pas réalisable numériquement sur ordinateur. Il est utile dans les méthodes de type faisceau qui résolvent le problème de la recherche d'un minimum d'une fonction convexe non différentiable.

On a en fait

$$0 \in \partial_\varepsilon f(\bar{x}) \Leftrightarrow f(\bar{x}) \leq f(x) + \varepsilon, \quad \forall x \in \mathbb{R}^n,$$

c'est-à-dire un zéro de $\partial_\varepsilon f$ est une solution à ε près de $\min f(x)$.

On va maintenant introduire un ε -élargissement pour les opérateurs maximaux monotones généraux. On va montrer que cet ε -élargissement a des propriétés semblables à celles du ε -sous-différentiel et on va l'utiliser pour développer des méthodes de recherche d'un zéro d'un opérateur maximal monotone.

Dans la suite, si on ne précise rien, on va supposer sans perte de généralité que T est défini sur l'espace entier, c'est-à-dire $D(T) = X$.

1.2 ε -élargissement d'un opérateur maximal monotone

Définition 1.10

Soient X un espace de Hilbert,

$T : X \rightarrow \mathcal{P}(X)$ un opérateur maximal monotone, et

$\varepsilon \geq 0$.

L' ε -élargissement de T en x est défini par

$$T^\varepsilon : X \rightarrow \mathcal{P}(X),$$

$$x \mapsto \{u \in X \mid \langle v - u, y - x \rangle \geq -\varepsilon, \quad \forall y \in X, \forall v \in T(y)\}.$$

En général, d'après la définition, T^ε n'est pas monotone, de même que $\partial_\varepsilon f$. Les deux opérateurs sont reliés par la proposition suivante :

Proposition 1.3

Soit $f : \mathbb{R}^n \rightarrow]-\infty, +\infty]$ une fonction convexe.

Si $T = \partial f$, alors $\partial_\varepsilon f \subseteq T^\varepsilon$.

Preuve : Soient $x \in \mathbb{R}^n$, $u \in \partial_\varepsilon f(x)$ et $y \in \mathbb{R}^n$ quelconques.
On a alors, par définition du ε -sous-différentiel

$$f(y) - f(x) \geq \langle u, y - x \rangle - \varepsilon .$$

Dès lors, pour tout $v \in T(y) = \partial f(y)$, on a

$$\varepsilon + \langle u, x - y \rangle \geq f(x) - f(y) \geq \langle v, x - y \rangle$$

et donc

$$\langle u - v, x - y \rangle \geq -\varepsilon , \quad \forall y \in \mathbb{R}^n , \forall v \in T(y) ,$$

c'est-à-dire $u \in T^\varepsilon(x)$. ■

Malheureusement, on n'a pas toujours $\partial_\varepsilon f = T^\varepsilon$. En effet, on peut avoir $\partial_\varepsilon f = T^\varepsilon$ pour un certain $\bar{\varepsilon} > 0$ différent de ε ou bien $\partial_\varepsilon f \subsetneq T^\varepsilon$ comme le montrent les exemples suivants.

Exemples :

1. Soient $f : \mathbb{R} \rightarrow \mathbb{R}$, $x \mapsto |x|$, et $T = \partial f$.

On trouve

$$\partial_\varepsilon f(x) = T^\varepsilon(x) = \begin{cases} \left[1 - \frac{\varepsilon}{x}, 1\right] , & \text{si } x > \frac{\varepsilon}{2} , \\ [-1, 1] , & \text{si } -\frac{\varepsilon}{2} \leq x \leq \frac{\varepsilon}{2} , \\ \left[-1, -1 + \frac{\varepsilon}{x}\right] , & \text{si } x < -\frac{\varepsilon}{2} . \end{cases}$$

2. Soient $f : \mathbb{R} \rightarrow \mathbb{R}$, $x \mapsto \frac{1}{2} x^2$, et $T = \partial f$.

On trouve

$$\partial_\varepsilon f(x) = [x - \sqrt{2\varepsilon}, x + \sqrt{2\varepsilon}] ,$$

$$T^\varepsilon(x) = [x - 2\sqrt{\varepsilon}, x + 2\sqrt{\varepsilon}] .$$

Dans ce cas, on a $\partial_{2\varepsilon} f = T^\varepsilon$.

3. Soient $f : \mathbb{R} \rightarrow \mathbb{R}$, $x \mapsto -\ln x$, et $T = \partial f$.

On trouve

$$\partial_\varepsilon f(x) = \left[-\frac{1}{x} \Delta_1(\varepsilon), -\frac{1}{x} \Delta_2(\varepsilon)\right] ,$$

$$T^\varepsilon(x) = \left[\frac{1}{x} (-1 - \varepsilon - 2\sqrt{\varepsilon}), \frac{1}{x} \min\{-1 + \varepsilon, 0\}\right]$$

où $0 < \Delta_2(\varepsilon) \leq 1 \leq \Delta_1(\varepsilon)$ sont les deux racines de $s - 1 - \ln s = \varepsilon$.

D'où $0 \in T^\varepsilon(x)$, $\forall x > 0$, $\forall \varepsilon \geq 1$,

mais $0 \notin \partial_\varepsilon f(x)$, ($\varepsilon \geq 0$).

Dans ce cas, on a donc $\partial_\varepsilon f \subsetneq T^\varepsilon$ pour $\varepsilon \geq 1$.

1.3 Propriétés de T^ε

Citons tout d'abord quelques propriétés élémentaires de T^ε qui s'établissent immédiatement à partir de la définition.

Proposition 1.4

Soient $T, T_1, T_2 : X \rightarrow \mathcal{P}(X)$.

$$1. \ 0 \leq \varepsilon_1 \leq \varepsilon_2 \quad \Rightarrow \quad T^{\varepsilon_1} \subseteq T^{\varepsilon_2}.$$

$$2. \ \forall \varepsilon_1, \varepsilon_2 \geq 0 : (T^{\varepsilon_1} + T^{\varepsilon_2}) \subseteq (T_1 + T_2)^{\varepsilon_1 + \varepsilon_2}.$$

$$3. \ \forall \alpha \geq 0, \forall \varepsilon \geq 0 : (\alpha T)^{\alpha \varepsilon} = \alpha T^\varepsilon.$$

$$4. \ \forall \alpha \in [0, 1], \forall \varepsilon \geq 0 : \alpha T_1^\varepsilon + (1 - \alpha) T_2^\varepsilon \subseteq (\alpha T_1 + (1 - \alpha) T_2)^\varepsilon.$$

$$5. \ \text{Si } E \subset \mathbb{R}^+, \text{ alors } \bar{\varepsilon} \stackrel{\text{not}}{=} \inf E \quad \Rightarrow \quad \bigcap_{\varepsilon \in E} T^\varepsilon = T^{\bar{\varepsilon}}.$$

$$6. \ \forall x \in X, \forall \varepsilon \geq 0 : T^\varepsilon(x) \text{ est convexe et fermé.}$$

Proposition 1.5

Le graphe de T^ε est fermé.

Autrement dit, pour toute suite $\{(\varepsilon_i, x^i, w^i \in T^{\varepsilon_i}(x_i))\}_{i \geq 0}$ telle que $\varepsilon_i \geq 0$, $\forall i$,

si

$$\lim_{i \rightarrow +\infty} x^i = x, \quad \lim_{i \rightarrow +\infty} w^i = w, \quad \lim_{i \rightarrow +\infty} \varepsilon_i = \varepsilon,$$

alors $w \in T^\varepsilon(x)$.

Preuve : Soient $y \in X$, $v \in T(y)$ quelconques.

On a $w^i \in T^{\varepsilon_i}(x_i)$ pour tout i , d'où, par définition de T^{ε_i} :

$$\langle v - w^i, y - x^i \rangle \geq -\varepsilon_i, \quad \forall i.$$

En passant à la limite pour $i \rightarrow +\infty$, on a, par continuité du produit scalaire :

$$\langle v - w, y - x \rangle \geq -\varepsilon$$

et donc, de nouveau par la Définition 1.10 de T^ε , $w \in T^\varepsilon(x)$. ■

Remarquons que ces deux propositions sont vraies même si T n'est pas monotone.

Si T est maximal monotone, on a en plus $T^0 = T$. En effet

$$T^0(x) = \{u \in X \mid \langle x - y, u - v \rangle \geq 0, \quad \forall y \in X, \forall v \in T(y)\}.$$

Si T est monotone, cette relation est toujours vérifiée et elle caractérise T entièrement s'il est maximal.

On va maintenant établir d'autres propriétés de T^ε , semblables à celles de $\partial_\varepsilon f$, qui seront utiles dans les preuves de convergence des algorithmes qu'on développera pour chercher un zéro d'un opérateur maximal monotone.

Définition 1.11

Un opérateur T sur \mathbb{R}^n est **localement borné** si, pour tout $x \in \mathring{D}(T)$, il existe un rayon $\delta_x > 0$, tel que $T(B(x, \delta_x))$ soit borné dans \mathbb{R}^n .

Définition 1.12

Un opérateur T sur \mathbb{R}^n est **borné sur des parties bornées** si et seulement si, pour toute partie bornée Q de \mathbb{R}^n telle que $\overline{Q} \subset \mathring{D}(T)$, l'ensemble $T(Q)$ est borné.

Proposition 1.6

Tout opérateur localement borné est borné sur des parties bornées.

Preuve : voir [2]. ■

Proposition 1.7

Si T est un opérateur maximal monotone, alors T est localement borné.

Preuve : voir [10]. ■

Théorème de Brønsted et Rockafellar

Le théorème dit que, pour une fonction convexe propre, fermée, chaque ε -sous-gradient en un point y peut être approché par un sous-gradient exact calculé en un point x (qui peut être différent de y).

Plus précisément,

Soient f une fonction convexe propre, fermée, $x \in \text{dom } f$ et $\varepsilon \geq 0$.

Alors, pour tout $\eta > 0$ et $s \in \partial_\varepsilon f(x)$, il existe $x_\eta \in B(x, \eta)$ et $s_\eta \in \partial f(x_\eta)$ tels que $\|s_\eta - s\| \leq \frac{\varepsilon}{\eta}$.

Sous forme ensembliste,

$$\partial_\varepsilon f(x) \subset \bigcap_{\eta > 0} \bigcup_{\|y-x\| \leq \eta} \left\{ \partial f(y) + B\left(0, \frac{\varepsilon}{\eta}\right) \right\}.$$

Ainsi, si y est proche de x (η petit), il se peut qu'on ait une mauvaise approximation. Si, par contre, l'approximation est bonne (η grand), il se peut que y soit loin de x .

Prendre $\eta = \sqrt{\varepsilon}$ est un compromis qui égale la distance à x et le degré d'approximation.

Pour l' ε -élargissement d'un opérateur maximal monotone, on a :

Théorème 1.2

Soient $T : X \rightarrow \mathcal{P}(X)$ un opérateur maximal monotone, $\varepsilon > 0$ et $(x_\varepsilon, v_\varepsilon) \in G(T^\varepsilon)$ ou

$$G(T^\varepsilon) = \{(x_\varepsilon, v_\varepsilon) \mid v_\varepsilon \in T^\varepsilon(x_\varepsilon)\}.$$

Alors, pour tout $\eta > 0$, il existe $(x, v) \in G(T)$ telle que $\|v - v_\varepsilon\| \leq \frac{\varepsilon}{\eta}$ et $\|x - x_\varepsilon\| \leq \eta$.

Sous forme ensembliste :

$$T^\varepsilon(x) \subset \bigcap_{\eta > 0} \bigcup_{\|y-x\| \leq \eta} \left\{ T(y) + B\left(0, \frac{\varepsilon}{\eta}\right) \right\}.$$

Preuve : Soit $\beta > 0$ quelconque.

On définit la fonction multivoque :

$$\begin{aligned} G_\beta : X &\rightarrow \mathcal{P}(X), \\ y &\mapsto \beta \cdot T(y) + \{y\}, \end{aligned}$$

c'est-à-dire

$$G_\beta = \beta \cdot T + I.$$

Comme T est maximal monotone et $\beta > 0$, βT est aussi maximal monotone et donc, par l'équivalence 1. \Leftrightarrow 2. du Théorème 1.1, on a

$$R(\beta T + I) = R(G_\beta) = X.$$

D'où G_β est surjectif et il existe $(x, v) \in G(T)$ telle que

$$\beta v_\varepsilon + x_\varepsilon = G_\beta(x) = \beta \cdot v + x. \quad (1.3)$$

Comme, par hypothèse, $(x_\varepsilon, v_\varepsilon) \in G(T^\varepsilon)$, il suit de la définition de $G(T^\varepsilon)$ et T^ε que

$$\langle w - v_\varepsilon, y - x_\varepsilon \rangle \geq -\varepsilon, \quad \forall y \in X, \forall w \in T(y).$$

Donc, en particulier pour $y = x$ et $w = v \in T(x)$, on a

$$\langle v - v_\varepsilon, x - x_\varepsilon \rangle \geq -\varepsilon.$$

En tirant x de (1.3), on en déduit les inégalités suivantes :

$$\begin{aligned} \langle v - v_\varepsilon, -\beta \cdot v + \beta \cdot v_\varepsilon + x_\varepsilon - x_\varepsilon \rangle &\geq -\varepsilon \\ \langle v - v_\varepsilon, \beta (v_\varepsilon - v) \rangle &\geq -\varepsilon \\ -\beta \langle v - v_\varepsilon, v - v_\varepsilon \rangle &\geq -\varepsilon \\ \|v - v_\varepsilon\|^2 &\leq \frac{\varepsilon}{\beta} \quad (\beta > 0). \end{aligned}$$

De même, en tirant v de (1.3), on a successivement :

$$\begin{aligned} \left\langle \frac{1}{\beta} (\beta v_\varepsilon + x_\varepsilon - x) - v_\varepsilon, x - x_\varepsilon \right\rangle &\geq -\varepsilon \\ \left\langle \frac{1}{\beta} (x_\varepsilon - x), x - x_\varepsilon \right\rangle &\geq -\varepsilon \\ -\frac{1}{\beta} \langle x - x_\varepsilon, x - x_\varepsilon \rangle &\geq -\varepsilon \\ \|x - x_\varepsilon\|^2 &\leq \beta \varepsilon \quad (\beta > 0). \end{aligned}$$

Le résultat suit alors en choisissant $\beta = \frac{\eta^2}{\varepsilon}$; en effet,

$$\|v - v_\varepsilon\|^2 \leq \frac{\varepsilon}{\beta} = \varepsilon \cdot \frac{\varepsilon}{\eta^2} = \frac{\varepsilon^2}{\eta^2} \quad \Rightarrow \quad \|v - v_\varepsilon\| \leq \frac{\varepsilon}{\eta}$$

et

$$\|x - x_\varepsilon\|^2 \leq \beta \cdot \varepsilon = \frac{\eta^2}{\varepsilon} \cdot \varepsilon = \eta^2 \quad \Rightarrow \quad \|x - x_\varepsilon\| \leq \eta . \quad \blacksquare$$

Remarques :

1. Dans la preuve, on a uniquement utilisé l' ε -inégalité de la définition; le résultat est donc valable pour tout autre élargissement pourvu qu'il soit contenu dans T^ε .
2. Prendre $\eta = \sqrt{\varepsilon}$ constitue de nouveau un compromis entre la distance à x et le degré d'approximation.

On peut alors écrire

$$T^\varepsilon(x) \subset \bigcup_{y \in B(x, \sqrt{\varepsilon})} \{T(y) + B(0, \sqrt{\varepsilon})\} .$$

Ainsi, T^ε est localement borné si T l'est.

En fait, ces deux opérateurs sont étroitement reliés par le résultat suivant :

Corollaire 1.1

Soit $T : X \rightarrow \mathcal{P}(X)$ un opérateur maximal monotone. Alors

1. $R(T) \subset R(T^\varepsilon) \subset \overline{R(T)}$.
2. $D(T) \subset D(T^\varepsilon) \subset \overline{D(T)}$.
3. $d((x_\varepsilon, v_\varepsilon), G(T)) \leq \sqrt{2\varepsilon}$.

Preuve :

1. On a

$$u \in R(T) \quad \Leftrightarrow \quad \exists x \text{ tel que } u \in T(x) .$$

Comme T est monotone

$$\langle u - v, x - y \rangle \geq 0$$

donc, pour $\varepsilon > 0$,

$$\langle u - v, x - y \rangle \geq -\varepsilon, \quad \forall y \in X, \forall v \in T(y)$$

d'où $u \in T^\varepsilon(x)$, c'est-à-dire $u \in R(T^\varepsilon)$ et donc $R(T) \subset R(T^\varepsilon)$.

Pour montrer la deuxième inclusion, appliquons le théorème 1.2 avec $\eta = 1, 2, 3, \dots$

On obtient ainsi une suite $(x_\eta, v_\eta) \in G(T)$ telle que

$$\|v_\eta - v_\varepsilon\| \leq \frac{\varepsilon}{\eta} \quad \text{et} \quad \|x_\eta - x_\varepsilon\| \leq \eta.$$

En faisant tendre $\eta \rightarrow +\infty$, on a donc une suite $(v_\eta) \subset R(T)$ telle que $\|v_\eta - v_\varepsilon\| \rightarrow 0$, c'est-à-dire

$$v_\eta \rightarrow v_\varepsilon \in T(x_\varepsilon)$$

d'où le point $v_\varepsilon \in R(T^\varepsilon)$ est un point d'adhérence d'une suite dans $R(T)$.

Par conséquent, $v_\varepsilon \in \overline{R(T)}$ et donc $R(T^\varepsilon) \subset \overline{R(T)}$.

2. On a

$$\begin{aligned} x \in D(T) &\Leftrightarrow T(x) \neq \emptyset \\ &\Leftrightarrow \exists u \in T(x) \\ &\Rightarrow u \in T^\varepsilon(x) && [\text{cf. 1., } R(T) \subset R(T^\varepsilon)] \\ &\Rightarrow T^\varepsilon(x) \neq \emptyset \\ &\Leftrightarrow x \in D(T^\varepsilon) \end{aligned}$$

d'où

$$D(T) \subset D(T^\varepsilon).$$

Pour montrer la deuxième inclusion, appliquons le théorème 1.2 avec $\eta = 1, \frac{1}{2}, \frac{1}{3}, \dots$

On obtient ainsi une suite $(x_\eta, v_\eta) \in G(T)$ telle que

$$\|v_\eta - v_\varepsilon\| \leq \frac{\varepsilon}{\eta} \quad \text{et} \quad \|x_\eta - x_\varepsilon\| \leq \eta.$$

En faisant tendre $\eta \rightarrow 0$, on a donc une suite $(x_\eta) \subset D(T)$ telle que $\|x_\eta - x_\varepsilon\| \rightarrow 0$, c'est-à-dire

$$x_\eta \rightarrow x_\varepsilon \in D(T^\varepsilon)$$

d'où le point $x_\varepsilon \in D(T^\varepsilon)$ est point d'adhérence d'une suite dans $D(T)$.

Par conséquent, $x_\varepsilon \in \overline{D(T)}$ et donc $D(T^\varepsilon) \subset \overline{D(T)}$.

3. On a immédiatement

$$\begin{aligned} d((x_\varepsilon, v_\varepsilon), G(T))^2 &= \min_{(y,u) \in G(T)} d((x_\varepsilon, v_\varepsilon), (y, u))^2 \\ &\leq d((x_\varepsilon, v_\varepsilon), (x, v))^2 \\ &\leq \|x - x_\varepsilon\|^2 + \|v - v_\varepsilon\|^2 \end{aligned}$$

et, par le théorème 1.2 appliqué avec $\eta = \sqrt{\varepsilon}$,

$$\begin{aligned} d((x_\varepsilon, v_\varepsilon), G(T))^2 &\leq (\sqrt{\varepsilon})^2 + \left(\frac{\varepsilon}{\sqrt{\varepsilon}}\right)^2 \\ &= \varepsilon + \frac{\varepsilon^2}{\varepsilon} \\ &= 2\varepsilon. \end{aligned}$$

Dès lors, $d((x_\varepsilon, v_\varepsilon), G(T)) \leq \sqrt{2\varepsilon}$. ■

Formule de Transport

On a déjà mentionné que l' ε -sous-différentiel est une approximation du sous-différentiel. Cependant, celle-ci ne sert à rien si on ne sait pas déterminer des éléments de $\partial_\varepsilon f$.

La question est alors la suivante: étant donné un sous-gradient exact s' calculé en un point x' , comment peut-on voir si ce s' appartient à l'ensemble $\partial_\varepsilon f(x)$ auquel on s'intéresse?

La réponse est donnée par le théorème suivant:

Théorème 1.3

Soient $x, x' \in \text{dom } f$ et $s' \in \partial f(x')$.

Alors $s' \in \partial_\varepsilon f(x)$, si et seulement si $\varepsilon \geq f(x) - f(x') - \langle s', x - x' \rangle$.

Preuve :

– La condition est suffisante.

En effet, comme $s' \in \partial_\varepsilon f(x)$, on a par définition

$$\forall y \in \mathbb{R}^n : \quad f(y) \geq f(x) + \langle s', y - x \rangle - \varepsilon .$$

Donc, en particulier pour $y = x'$,

$$f(x') \geq f(x) + \langle s', x' - x \rangle - \varepsilon .$$

– La condition est nécessaire.

Soit $y \in \mathbb{R}^n$ arbitraire.

On a, par hypothèse,

$$s' \in \partial f(x') \quad \text{et} \quad f(x') \geq f(x) + \langle s', x' - x \rangle - \varepsilon .$$

On obtient ainsi successivement

$$\begin{aligned} f(y) &\geq f(x') + \langle s', y - x' \rangle , \\ f(y) &\geq f(x) + \langle s', x' - x \rangle - \varepsilon + \langle s', y - x' \rangle , \\ f(y) &\geq f(x) + \langle s', y - x \rangle - \varepsilon . \end{aligned}$$

Comme $y \in \mathbb{R}^n$ est arbitraire, on en déduit que $s' \in \partial_\varepsilon f(x)$. ■

On appelle ce théorème **formule de transport** car il “transporte” en x un sous-gradient donné en x' .

Pour l' ε -élargissement d'un opérateur maximal monotone, on veut aussi construire un élément de $T^\varepsilon(x)$ à l'aide de quelques éléments $(z^i, w^i) \in G(T)$. Mais la formule qu'on obtient est plus faible que celle pour le sous-différentiel dans le sens qu'ici, on peut “transporter” uniquement des sommes convexes.

Théorème 1.4

Soit $T : X \rightarrow \mathcal{P}(X)$ un opérateur maximal monotone.

Considérons un ensemble de m triplets

$$\{(\varepsilon_i, z^i, w^i)\}_{i=1, \dots, m}$$

où, pour tout i , $\varepsilon_i \geq 0$, $z^i \in X$ et $w^i \in T^{\varepsilon_i}(z^i)$.

Pour chaque $\alpha \in \Delta_m \stackrel{\text{def}}{=} \{\alpha \in \mathbb{R}^m \mid \alpha_i \geq 0, \sum_{i=1}^m \alpha_i = 1\}$, définissons

$$\hat{z} = \sum_{i=1}^m \alpha_i z^i,$$

$$\hat{w} = \sum_{i=1}^m \alpha_i w^i$$

$$\hat{\varepsilon} = \sum_{i=1}^m \alpha_i \varepsilon_i + \sum_{i=1}^m \alpha_i \langle z^i - \hat{z}, w^i - \hat{w} \rangle.$$

Alors $\hat{\varepsilon} \geq 0$ et $\hat{w} \in T^{\hat{\varepsilon}}(\hat{z})$.

Preuve : Il faut montrer que $\hat{w} \in T^{\hat{\varepsilon}}(\hat{z})$, c'est-à-dire que

$$\langle v - \hat{w}, y - \hat{z} \rangle \geq -\hat{\varepsilon}, \quad \forall (y, v) \in G(T).$$

Par définition de \hat{z} , pour tout $(y, v) \in G(T)$, on a :

$$\begin{aligned} \langle z - y, \hat{w} - v \rangle &= \sum_{i=1}^m \alpha_i \langle z^i - y, \hat{w} - v \rangle \\ &= \sum_{i=1}^m \alpha_i \langle z^i - y, \hat{w} - w^i + w^i - v \rangle \\ &= \sum_{i=1}^m \alpha_i [\langle z^i - y, \hat{w} - w^i \rangle + \langle z^i - y, w^i - v \rangle] \\ &= \sum_{i=1}^m \alpha_i \langle z^i - y, \hat{w} - w^i \rangle + \sum_{i=1}^m \alpha_i \langle z^i - y, w^i - v \rangle. \end{aligned}$$

Comme $w^i \in T^{\varepsilon_i}(z^i)$, on a $\langle z^i - y, w^i - v \rangle \geq -\varepsilon_i$

et donc

$$\langle \hat{z} - y, \hat{w} - v \rangle \geq \sum_{i=1}^m \alpha_i \langle z^i - y, \hat{w} - w^i \rangle + \sum_{i=1}^m \alpha_i (-\varepsilon_i). \quad (1.4)$$

D'autre part, comme $\sum_{i=1}^m \alpha_i = 1$ et $\hat{w} = \sum_{i=1}^m \alpha_i w^i$, on obtient

$$\begin{aligned}
\sum_{i=1}^m \alpha_i \langle z^i - y, \hat{w} - w^i \rangle &= \sum_{i=1}^m \alpha_i \langle z^i - \hat{z} + \hat{z} - y, \hat{w} - w^i \rangle \\
&= \sum_{i=1}^m \alpha_i [\langle z^i - \hat{z}, \hat{w} - w^i \rangle + \langle \hat{z} - y, \hat{w} - w^i \rangle] \\
&= - \sum_{i=1}^m \alpha_i \langle z^i - \hat{z}, w^i - \hat{w} \rangle + \langle \hat{z} - y, \sum_{i=1}^m \alpha_i (\hat{w} - w^i) \rangle \\
&= - \sum_{i=1}^m \alpha_i \langle z^i - \hat{z}, w^i - \hat{w} \rangle + \langle \hat{z} - y, \hat{w} - \sum_{i=1}^m \alpha_i w^i \rangle \\
&= - \sum_{i=1}^m \alpha_i \langle z^i - \hat{z}, w^i - \hat{w} \rangle
\end{aligned}$$

d'où l'inégalité (1.4) entraîne

$$\langle \hat{z} - y, \hat{w} - v \rangle \geq - \left[\sum_{i=1}^m \alpha_i \varepsilon_i + \sum_{i=1}^m \alpha_i \langle z^i - \hat{z}, w^i - \hat{w} \rangle \right] = -\hat{\varepsilon} .$$

Montrons que $\hat{\varepsilon} \geq 0$. Pour cela, supposons par l'absurde que $\hat{\varepsilon} < 0$.

Alors

$$\langle \hat{z} - y, \hat{w} - v \rangle \geq -\hat{\varepsilon} > 0 , \quad \forall (y, v) \in G(T)$$

implique que $\hat{w} \in T(\hat{z})$, c'est-à-dire $(\hat{z}, \hat{w}) \in G(T)$ car T est maximal [cf. définition 1.6].

Pour $(y, v) = (\hat{z}, \hat{w}) \in G(T)$, on a alors

$$\langle \hat{z} - \hat{z}, \hat{w} - \hat{w} \rangle > 0$$

c'est-à-dire $0 > 0$, ce qui est impossible.

D'où $\hat{\varepsilon} \geq 0$.

Finalement, on a

$$\langle \hat{z} - y, \hat{w} - v \rangle \geq \hat{\varepsilon} , \quad \forall (y, v) \in G(T) , \quad \hat{\varepsilon} \geq 0 ,$$

et donc, par définition, que $\hat{w} \in T^{\hat{\varepsilon}}(\hat{z})$. ■

La formule de transport pour l' ε -sous-différentiel (Théorème 1.3) peut également être utilisée pour obtenir la borne inférieure suivante :

$$\langle x^1 - x^2, v^1 - v^2 \rangle \geq -(\varepsilon_1 + \varepsilon_2)$$

pour tout $v^i \in \partial_{\varepsilon_i} f(x^i)$, $i = 1, 2$.

Dans le cas général, ce résultat devient

Corollaire 1.2

Soient $w^1 \in T^{\varepsilon_1}(z^1)$ et $w^2 \in T^{\varepsilon_2}(z^2)$.

Alors

$$\langle z^1 - z^2, w^1 - w^2 \rangle \geq -(\sqrt{\varepsilon_1} + \sqrt{\varepsilon_2})^2.$$

Preuve : Distinguons les trois cas suivants :

1er cas : $\varepsilon_1 = 0$ et $\varepsilon_2 = 0$

Dans ce cas, $w^1 \in T(z^1)$, $w^2 \in T(z^2)$ et, comme T est monotone, on a immédiatement la thèse

$$\langle z^1 - z^2, w^1 - w^2 \rangle \geq 0.$$

2ème cas : $\varepsilon_1 \neq 0$ et $\varepsilon_2 = 0$

(le raisonnement est semblable pour $\varepsilon_1 = 0$ et $\varepsilon_2 \neq 0$).

Dans ce cas, $w^1 \in T^{\varepsilon_1}(z^1)$ et $w^2 \in T(z^2)$.

Dès lors par définition de T^{ε_1} , on a

$$\langle v - w^1, y - z^1 \rangle \geq -\varepsilon_1, \quad \forall y \in X, \forall v \in T(y)$$

et en particulier, pour $y = z^2$, $v = w^2 \in T(z^2)$,

$$\langle w^2 - w^1, z^2 - z^1 \rangle \geq -\varepsilon_1$$

c'est-à-dire la thèse

$$\langle z^1 - z^2, w^1 - w^2 \rangle \geq -(\sqrt{\varepsilon_1})^2.$$

3ème cas : $\varepsilon_1 \neq 0$ et $\varepsilon_2 \neq 0$

Choisissons

$$\alpha_1 = \frac{\sqrt{\varepsilon_2}}{\sqrt{\varepsilon_1} + \sqrt{\varepsilon_2}} \quad \text{et} \quad \alpha_2 = 1 - \alpha_1 = \frac{\sqrt{\varepsilon_1}}{\sqrt{\varepsilon_1} + \sqrt{\varepsilon_2}},$$

et définissons

$$\begin{cases} \hat{z} &= \alpha_1 z^1 + \alpha_2 z^2, \\ \hat{w} &= \alpha_1 w^1 + \alpha_2 w^2, \\ \hat{\varepsilon} &= \alpha_1 \varepsilon_1 + \alpha_2 \varepsilon_2 + \alpha_1 \langle z^1 - \hat{z}, w^1 - \hat{w} \rangle + \alpha_2 \langle z^2 - \hat{z}, w^2 - \hat{w} \rangle. \end{cases}$$

Alors

$$\begin{aligned} z^1 - \hat{z} &= z^1 - \alpha_1 z^1 - \alpha_2 z^2 \\ &= (1 - \alpha_1) z^1 - \alpha_2 z^2 \\ &= \alpha_2 z^1 - \alpha_2 z^2 \\ &= \alpha_2 (z^1 - z^2). \end{aligned}$$

De même, nous pouvons écrire les égalités suivantes

$$\begin{aligned} z^2 - \hat{z} &= z^2 - \alpha_1 z^1 - \alpha_2 z^2 \\ &= (1 - \alpha_2) z^2 - \alpha_1 z^1 \\ &= \alpha_1 z^2 - \alpha_1 z^1 \\ &= \alpha_1 (z^2 - z^1), \end{aligned}$$

$$\begin{aligned} w^1 - \hat{w} &= w^1 - \alpha_1 w^1 - \alpha_2 w^2 \\ &= (1 - \alpha_1) w^1 - \alpha_2 w^2 \\ &= \alpha_2 w^1 - \alpha_2 w^2 \\ &= \alpha_2 (w^1 - w^2), \end{aligned}$$

$$\begin{aligned} w^2 - \hat{w} &= w^2 - \alpha_1 w^1 - \alpha_2 w^2 \\ &= (1 - \alpha_2) w^2 - \alpha_1 w^1 \\ &= \alpha_1 w^2 - \alpha_1 w^1 \\ &= \alpha_1 (w^2 - w^1). \end{aligned}$$

Avec ces développements, $\hat{\varepsilon}$ peut s'écrire comme suit :

$$\begin{aligned} \hat{\varepsilon} &= \alpha_1 \varepsilon_1 + \alpha_2 \varepsilon_2 + \alpha_1 \langle \alpha_2 (z^1 - z^2), \alpha_2 (w^1 - w^2) \rangle \\ &\quad + \alpha_2 \langle \alpha_1 (z^2 - z^1), \alpha_1 (w^2 - w^1) \rangle \\ &= \alpha_1 \varepsilon_1 + \alpha_2 \varepsilon_2 + \alpha_1 \alpha_2^2 \langle z^1 - z^2, w^1 - w^2 \rangle + \alpha_2 \alpha_1^2 \langle z^2 - z^1, w^2 - w^1 \rangle \\ &= \alpha_1 \varepsilon_1 + \alpha_2 \varepsilon_2 + [\alpha_1 \alpha_2^2 + \alpha_2 \alpha_1^2] [\langle z^2 - z^1, w^2 - w^1 \rangle]. \end{aligned}$$

Remplaçant ensuite les α_i par leur définition, on a successivement

$$\begin{aligned}
\hat{\varepsilon} &= \frac{\sqrt{\varepsilon_2}}{\sqrt{\varepsilon_1} + \sqrt{\varepsilon_2}} \cdot \varepsilon_1 + \frac{\sqrt{\varepsilon_1}}{\sqrt{\varepsilon_1} + \sqrt{\varepsilon_2}} \cdot \varepsilon_2 \\
&\quad + [\langle z^2 - z^1, w^2 - w^1 \rangle] \cdot \left[\frac{\sqrt{\varepsilon_2}}{\sqrt{\varepsilon_1} + \sqrt{\varepsilon_2}} \cdot \frac{\varepsilon_1}{(\sqrt{\varepsilon_1} + \sqrt{\varepsilon_2})^2} + \frac{\varepsilon_2}{(\sqrt{\varepsilon_1} + \sqrt{\varepsilon_2})^2} \cdot \frac{\sqrt{\varepsilon_1}}{\sqrt{\varepsilon_1} + \sqrt{\varepsilon_2}} \right] \\
&= \sqrt{\varepsilon_1} \sqrt{\varepsilon_2} \left[\frac{\sqrt{\varepsilon_1}}{\sqrt{\varepsilon_1} + \sqrt{\varepsilon_2}} + \frac{\sqrt{\varepsilon_2}}{\sqrt{\varepsilon_1} + \sqrt{\varepsilon_2}} \right] \\
&\quad + [\langle z^2 - z^1, w^2 - w^1 \rangle] \cdot \left[\frac{\sqrt{\varepsilon_1} \sqrt{\varepsilon_2} [\sqrt{\varepsilon_1} + \sqrt{\varepsilon_2}]}{(\sqrt{\varepsilon_1} + \sqrt{\varepsilon_2})^3} \right] \\
&= \sqrt{\varepsilon_1 \varepsilon_2} + [\langle z^2 - z^1, w^2 - w^1 \rangle] \cdot \frac{\sqrt{\varepsilon_1 \varepsilon_2}}{(\sqrt{\varepsilon_1} + \sqrt{\varepsilon_2})^2} .
\end{aligned}$$

Comme $\hat{\varepsilon} \geq 0$, par le Théorème 1.4, on en déduit que

$$\langle z^2 - z^1, w^2 - w^1 \rangle \frac{\sqrt{\varepsilon_1 \varepsilon_2}}{(\sqrt{\varepsilon_1} + \sqrt{\varepsilon_2})^2} \geq -\sqrt{\varepsilon_1 \varepsilon_2}$$

et donc la thèse

$$\langle z^1 - z^2, w^1 - w^2 \rangle \geq -(\sqrt{\varepsilon_1} + \sqrt{\varepsilon_2})^2 . \quad \blacksquare$$

Remarquons encore que, si $\varepsilon_i = 0$, $i = 1, \dots, m$, le Théorème 1.4 montre comment construire $\hat{w} \in T^{\varepsilon}(\hat{z})$ à partir de $(z^i, w^i) \in G(T)$.

De plus, ce théorème permet de borner $\hat{\varepsilon}$ uniformément :

Corollaire 1.3

Soient les hypothèses et notations du Théorème 1.4 et supposons en plus que $\varepsilon_i = 0$ pour tout i .

Soient $\tilde{x} \in X$ et $\rho > 0$ tels que $\|z^i - \tilde{x}\| \leq \rho$, $\forall i \leq m$.

Alors les sommes convexes

$$\begin{aligned}
\hat{z} &= \sum_{i=1}^m \alpha_i z^i , \\
\hat{w} &= \sum_{i=1}^m \alpha_i w^i ,
\end{aligned}$$

sont telles que

1. $\|\hat{z} - \tilde{x}\| \leq \rho$;
2. $\hat{w} \in T^{\hat{\varepsilon}}(\hat{z})$ avec

$$\hat{\varepsilon} = \sum_{i=1}^m \alpha_i \langle z^i - \hat{z}, w^i - \hat{w} \rangle \leq 2 \rho M$$

$$\text{où } M = \max\{\|w^i\| \mid i = 1, \dots, m\}.$$

Preuve : Comme $\sum_{i=1}^m \alpha_i = 1$, on a

$$\|\hat{z} - \tilde{x}\| = \left\| \sum_{i=1}^m \alpha_i z^i - \tilde{x} \right\| = \left\| \sum_{i=1}^m \alpha_i z^i - \sum_{i=1}^m \alpha_i \tilde{x} \right\| = \left| \sum_{i=1}^m \alpha_i \right| \|z^i - \tilde{x}\| = \|z^i - \tilde{x}\| \leq \rho$$

et

$$\|\hat{w}\| = \left\| \sum_{i=1}^m \alpha_i w^i \right\| = \left| \sum_{i=1}^m \alpha_i \right| \|w^i\| = \|w^i\| \leq M.$$

Par le Théorème 1.4 avec $\varepsilon_i = 0$ pour tout i , on a immédiatement que

$$\hat{w} \in T^{\hat{\varepsilon}}(\hat{z}) \quad \text{avec } \hat{\varepsilon} = \sum_{i=1}^m \alpha_i \langle z^i - \hat{z}, w^i - \hat{w} \rangle.$$

Il reste à montrer que $\hat{\varepsilon} \leq 2 \rho M$. Pour cela, remarquons que

$$\begin{aligned} & \sum_{i=1}^m \alpha_i \langle z^i - \hat{z}, w^i - \hat{w} \rangle \\ &= \sum_{i=1}^m \alpha_i \langle z^i - \hat{z} + \tilde{x} - \tilde{x}, w^i - \hat{w} \rangle \\ &= \sum_{i=1}^m \alpha_i \langle z^i - \tilde{x}, w^i - \hat{w} \rangle + \sum_{i=1}^m \alpha_i \langle \tilde{x} - \hat{z}, w^i - \hat{w} \rangle \\ &= \sum_{i=1}^m \alpha_i \langle z^i - \tilde{x}, w^i - \hat{w} \rangle + \left\langle \tilde{x} - \hat{z}, \sum_{i=1}^m \alpha_i w^i - \hat{w} \right\rangle \quad [\text{cf. } \sum_{i=1}^m \alpha_i = 1] \\ &= \sum_{i=1}^m \alpha_i \langle z^i - \tilde{x}, w^i - \hat{w} \rangle \quad [\text{cf. } \hat{w} = \sum_{i=1}^m \alpha_i w^i] \\ &\leq \sum_{i=1}^m \alpha_i \|z^i - \tilde{x}\| \|w^i - \hat{w}\| \quad [\text{Cauchy-Schwarz}] \\ &\leq \sum_{i=1}^m \alpha_i \rho (\|w^i\| + \|\hat{w}\|) \quad [\|z^i - \tilde{x}\| \leq \rho \text{ et inégalité triangulaire}] \end{aligned}$$

$$\begin{aligned}
&\leq \sum_{i=1}^m \alpha_i \rho \, 2 \, M && [\|w^i\| \leq M \text{ par définition, et } \|\hat{w}\| \leq M] \\
&= 2 \rho \, M . && [\text{cf. } \sum_{i=1}^m \alpha_i = 1]
\end{aligned}$$

D'où $\hat{\varepsilon} \leq 2 \rho \, M$. ■

Il reste à montrer une dernière propriété très importante de l' ε -élargissement :

L'application $(\varepsilon, x) \mapsto T^\varepsilon(x)$ est localement Lipschitz continue sur $\overset{\circ}{D}(T)$.

Dans ce paragraphe ainsi qu'aux Chapitres 2 et 3, on prend $X = \mathbb{R}^n$ car on veut que les ensembles fermés bornés soient compacts.

Montrons d'abord un lemme technique :

Lemme 1.2

Soient $K \subset \overset{\circ}{D}(T)$ un ensemble compact et $\rho > 0$ tels que

$$\widetilde{K} \equiv K + \overline{B(0, \rho)} \subset \overset{\circ}{D}(T) .$$

Définissons

$$\widetilde{M} = \sup\{\|u\| \mid u \in T(\widetilde{K})\} .$$

Alors,

$$\forall \varepsilon > 0 : \quad \sup\{\|u\| \mid u \in T^\varepsilon(K)\} \leq \frac{\varepsilon}{\rho} + \widetilde{M} .$$

Preuve : Comme K est compact et $\overline{B(0, \rho)}$ est fermée et bornée dans \mathbb{R}^n donc compacte, alors \widetilde{K} est compact.

Comme T est maximal monotone, par la Proposition 1.7, T est localement borné sur $\overset{\circ}{D}(T)$ et donc, par la Proposition 1.6, T est borné sur les parties bornées, ce qui signifie que $T(\widetilde{K})$ est borné et, par conséquent, \widetilde{M} est fini.

Soient $\varepsilon > 0$, $x_\varepsilon \in K$ et $v_\varepsilon \in T^\varepsilon(x_\varepsilon)$.

Par le Théorème 1.2 avec $\eta = \rho > 0$, on sait qu'il existe $(x, v) \in G(T)$ telle que

$$\|v - v_\varepsilon\| \leq \frac{\varepsilon}{\rho} \tag{1.5}$$

et

$$\|x - x_\varepsilon\| \leq \rho \quad \Leftrightarrow \quad x \in \overline{B(x_\varepsilon, \rho)} .$$

Comme $x_\varepsilon \in K$, on a

$$x \in K + \overline{B(0, \rho)} \equiv \widetilde{K} ;$$

d'où

$$v = T(x) \in T(\widetilde{K})$$

et, par conséquent,

$$\|v\| \leq \widetilde{M} . \quad (1.6)$$

Ainsi

$$\begin{aligned} \|v_\varepsilon\| &= \|v_\varepsilon - v + v\| \\ &\leq \|v_\varepsilon - v\| + \|v\| && \text{[inégalité triangulaire]} \\ &\leq \frac{\varepsilon}{\rho} + \widetilde{M} && \text{[cf. (1.5) et (1.6)]} \end{aligned}$$

et comme $\varepsilon > 0$ et $v_\varepsilon \in T^\varepsilon(K)$ étaient choisis arbitrairement, il s'ensuit :

$$\forall \varepsilon > 0 : \quad \sup\{\|u\| \mid u \in T^\varepsilon(K)\} \leq \frac{\varepsilon}{\rho} + \widetilde{M} . \quad \blacksquare$$

Remarquons que ce lemme permet également d'affirmer que

*Si T est maximal monotone et si $D(T)$ est fermée,
alors T^ε est borné sur les parties bornées.*

On peut maintenant prouver la Lipschitz continuité de T^ε .

Théorème 1.5

Soient $K \subset \overset{\circ}{D}(T)$ un ensemble compact et $0 < \underline{\varepsilon} \leq \bar{\varepsilon} < +\infty$.

Alors il existe deux constantes $A \geq 0$ et $B \geq 0$ telles que

$\forall (\varepsilon_1, x^1), (\varepsilon_2, x^2) \in [\underline{\varepsilon}, \bar{\varepsilon}] \times K$ et $v^1 \in T^{\varepsilon_1}(x^1)$:

$\exists v^2 \in T^{\varepsilon_2}(x^2)$ tel que

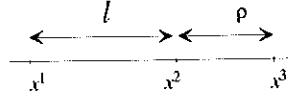
$$\|v^1 - v^2\| \leq A \|x^1 - x^2\| + B |\varepsilon^1 - \varepsilon^2| .$$

Preuve : Avec le ρ , \widetilde{K} et \widetilde{M} du Lemme 1.2, il suffit de prendre

$$A = \left(\frac{1}{\rho} + \frac{2 \widetilde{M}}{\underline{\varepsilon}} \right) \left(\frac{\bar{\varepsilon}}{\rho} + 2 \widetilde{M} \right), \quad B = \left(\frac{1}{\rho} + \frac{2 \widetilde{M}}{\underline{\varepsilon}} \right).$$

En effet, soient $\ell \equiv \|x^1 - x^2\|$ et x^3 aligné avec x^1 et x^2 tels que

$$\|x^3 - x^2\| = \rho \quad \text{et} \quad \|x^3 - x^1\| = \rho + \ell$$



c'est-à-dire $x^3 \in \overline{B(x^2, \rho)}$.

Comme $x^2 \in K$, on a

$$x^3 \in K + \overline{B(0, \rho)} \equiv \widetilde{K}.$$

Alors

$$x^2 = (1 - \theta) x^1 + \theta x^3 \quad \text{avec} \quad \theta = \frac{\ell}{\rho + \ell} \in [0, 1[.$$

Soit $u^3 \in T(x^3)$ et définissons

$$\tilde{v}^2 \equiv (1 - \theta) v^1 + \theta u^3.$$

Appliquons alors le Théorème 1.4 (formule de transport) à

$$\{(\varepsilon_1, x^1, v^1 \in T^{\varepsilon_1}(x^1)), (0, x^3, u^3 \in T(x^3))\},$$

$$\alpha_1 = 1 - \theta, \quad \alpha_2 = \theta;$$

$$\hat{z} = \alpha_1 z^1 + \alpha_2 z^2 = (1 - \theta) x^1 + \theta x^3 = x^2,$$

$$\hat{w} = \alpha_1 w^1 + \alpha_2 w^2 = (1 - \theta) v^1 + \theta u^3 \equiv \tilde{v}^2,$$

$$\begin{aligned} \hat{\varepsilon} &= \alpha_1 \varepsilon_1 + \alpha_2 \varepsilon_2 + \alpha_1 \langle z^1 - \hat{z}, w^1 - \hat{w} \rangle + \alpha_2 \langle z^2 - \hat{z}, w^2 - \hat{w} \rangle \\ &= (1 - \theta) \varepsilon_1 + \theta 0 + (1 - \theta) \langle x^1 - x^2, v^1 - \tilde{v}^2 \rangle + \theta \langle x^3 - x^2, u^3 - \tilde{v}^2 \rangle. \end{aligned}$$

En remplaçant x^2 et \tilde{v}^2 par leur définition,

$$\begin{aligned}
\hat{\varepsilon} &= (1 - \theta) \varepsilon_1 + (1 - \theta) \langle x^1 - (1 - \theta) x^1 - \theta x^3, v^1 - (1 - \theta) v^1 - \theta u^3 \rangle \\
&\quad + \theta \langle x^3 - (1 - \theta) x^1 - \theta x^3, u^3 - (1 - \theta) v^1 - \theta u^3 \rangle \\
&= (1 - \theta) \varepsilon_1 + (1 - \theta) \langle \theta x^1 - \theta x^3, \theta v^1 - \theta u^3 \rangle \\
&\quad + \theta \langle (1 - \theta) (x^3 - x^1), (1 - \theta) (u^3 - v^1) \rangle \\
&= (1 - \theta) \varepsilon_1 + (1 - \theta) \theta^2 \langle x^1 - x^3, v^1 - u^3 \rangle + \theta (1 - \theta)^2 \langle x^3 - x^1, u^3 - v^1 \rangle \\
&= (1 - \theta) \varepsilon_1 + \langle x^1 - x^3, v^1 - u^3 \rangle [\theta^2 - \theta^3 + \theta - 2\theta^2 + \theta^3] \\
&= (1 - \theta) \varepsilon_1 + \langle x^1 - x^3, v^1 - u^3 \rangle \theta (1 - \theta) \\
&\stackrel{\text{not}}{\equiv} \tilde{\varepsilon}_2 .
\end{aligned}$$

Par le théorème, $\hat{w} \in T^{\hat{\varepsilon}}(\hat{z})$, c'est-à-dire

$$\tilde{v}^2 \in T^{\tilde{\varepsilon}_2}(x^2) . \quad (1.7)$$

Comme $v^1 \in T^{\varepsilon_1}(x^1)$ et $u^3 \in T(x^3)$, en appliquant successivement l'inégalité triangulaire et le Lemme 1.2, on a :

$$\begin{aligned}
\|v^1 - u^3\| &\leq \|v^1\| + \|u^3\| \\
&\leq \left(\frac{\varepsilon_1}{\rho} + \tilde{M} \right) + \tilde{M} \\
&= \frac{\varepsilon_1}{\rho} + 2 \tilde{M} .
\end{aligned} \quad (1.8)$$

On obtient ainsi

$$\begin{aligned}
\tilde{\varepsilon}^2 &= (1 - \theta) \varepsilon_1 + \theta (1 - \theta) \langle x^1 - x^3, v^1 - u^3 \rangle \\
&\leq (1 - \theta) \varepsilon_1 + \theta (1 - \theta) \|x^1 - x^3\| \|v^1 - u^3\| && \text{[Cauchy-Schwarz]} \\
&\leq (1 - \theta) \varepsilon_1 + \theta (1 - \theta) (\ell + \rho) \left(\frac{\varepsilon_1}{\rho} + 2 \tilde{M} \right) && \text{[cf. } \|x^1 - x^3\| \equiv \ell + \rho \text{ , (1.8)]} \\
&= \frac{\rho + \ell - \ell}{\rho + \ell} \varepsilon_1 + \frac{\ell}{\rho + \ell} \frac{\rho}{\rho + \ell} (\rho + \ell) \left(\frac{\varepsilon_1}{\rho} + 2 \tilde{M} \right) && \text{[cf. } \theta \equiv \frac{\ell}{\rho + \ell} \text{]}
\end{aligned}$$

$$\begin{aligned}
&= \frac{\rho}{\rho + \ell} \varepsilon_1 + \frac{\ell}{\rho + \ell} \varepsilon_1 + \frac{\ell \rho}{\rho + \ell} 2 \widetilde{M} \\
&= \frac{\rho + \ell}{\rho + \ell} \varepsilon_1 + \frac{\rho \ell}{\rho + \ell} 2 \widetilde{M} \\
&= \varepsilon_1 + \frac{\rho \ell}{\rho + \ell} 2 \widetilde{M}
\end{aligned} \tag{1.9}$$

On a également

$$\begin{aligned}
\|v^1 - \tilde{v}^2\| &= \|v^1 - (1 - \theta) v^1 - \theta u^3\| && [\text{par définition de } \tilde{v}^2] \\
&= \|\theta v^1 - \theta u^3\| \\
&= \theta \|v^1 - u^3\| && [\theta \in [0, 1[] \\
&\leq \theta \left(\frac{\varepsilon_1}{\rho} + 2 \widetilde{M} \right) && (1.10) \\
&&& [\text{cf. (1.8)}]
\end{aligned}$$

de même que

$$\begin{aligned}
\|v^1 - \tilde{v}^2\| &\leq \frac{\ell}{\rho + \ell} \left(\frac{\varepsilon_1}{\rho} + 2 \widetilde{M} \right) && [\text{cf. } \theta \equiv \frac{\ell}{\rho + \ell}] \\
&\leq \frac{\ell}{\rho} \left(\frac{\varepsilon_1}{\rho} + 2 \widetilde{M} \right) && [\ell > 0] \\
&= \|x^1 - x^2\| \frac{1}{\rho} \left(\frac{\varepsilon_1}{\rho} + 2 \widetilde{M} \right) && (1.11) \\
&&& [\ell \equiv \|x^1 - x^2\|]
\end{aligned}$$

1er cas : $\tilde{\varepsilon}_2 \leq \varepsilon_2$

Par (1.7), on a

$$\tilde{v}^2 \in T^{\tilde{\varepsilon}_2}(x^2) \subseteq T^{\varepsilon_2}(x^2)$$

où la dernière inclusion provient du fait que $\tilde{\varepsilon}_2 \leq \varepsilon_2$ et de la Proposition 1.4, 1.

Si on choisit $v^2 \equiv \tilde{v}^2$, l'équation (1.11) donne

$$\begin{aligned} \|v^1 - v^2\| &\leq \|x^1 - x^2\| \frac{1}{\rho} \left(\frac{\varepsilon_1}{\rho} + 2 \widetilde{M} \right) \\ &\leq \|x^1 - x^2\| \left(\frac{1}{\rho} + \frac{2 \widetilde{M}}{\underline{\varepsilon}} \right) \left(\frac{\bar{\varepsilon}}{\rho} + 2 \widetilde{M} \right) + \left(\frac{1}{\rho} + \frac{2 \widetilde{M}}{\underline{\varepsilon}} \right) |\varepsilon^1 - \varepsilon^2| \\ &= A \|x^1 - x^2\| + B |\varepsilon^1 - \varepsilon^2| \end{aligned}$$

où la dernière inégalité tient car on n'a fait que rajouter des nombres positifs.

2ème cas : $\tilde{\varepsilon}_2 > \varepsilon_2$

Poser $\beta \equiv \frac{\varepsilon_2}{\tilde{\varepsilon}_2} < 1$ et choisir $v^2 = (1 - \beta) u^2 + \beta \tilde{v}^2$ avec $u^2 \in T(x^2)$.

Appliquons alors le Théorème 1.4 (formule de transport) à

$$\{(0, x^2, u^2 \in T(x^2)), (\tilde{\varepsilon}_2, x^2, \tilde{v}^2 \in T^{\tilde{\varepsilon}_2}(x^2))\} ;$$

$$\alpha_1 = 1 - \beta, \quad \alpha_2 = \beta ;$$

$$\hat{z} = \alpha_1 z^1 + \alpha_2 z^2 = (1 - \beta) x^2 + \beta x^2 = x^2 ,$$

$$\hat{w} = \alpha_1 w^1 + \alpha_2 w^2 = (1 - \beta) u^2 + \beta \tilde{v}^2 = v^2 \quad [\text{cf. choix}]$$

$$\begin{aligned} \hat{\varepsilon} &= \alpha_1 \varepsilon_1 + \alpha_2 \varepsilon_2 + \alpha_1 \langle z^1 - \hat{z}, w^1 - \hat{w} \rangle + \alpha_2 \langle z^2 - \hat{z}, w^2 - \hat{w} \rangle \\ &= (1 - \beta) 0 + \beta \tilde{\varepsilon}_2 + (1 - \beta) \langle x^2 - x^2, u^2 - v^2 \rangle + \beta \langle x^2 - x^2, \tilde{v}^2 - v^2 \rangle \\ &= \beta \tilde{\varepsilon}_2 + 0 + 0 \end{aligned}$$

$$= \frac{\varepsilon_2}{\tilde{\varepsilon}_2} \tilde{\varepsilon}_2 \quad [\text{cf. } \beta = \frac{\varepsilon_2}{\tilde{\varepsilon}_2}]$$

$$= \varepsilon_2 .$$

Par le théorème, $\hat{w} \in T^{\hat{\varepsilon}}(\hat{z})$, c'est-à-dire $v^2 \in T^{\varepsilon_2}(x^2)$.

De plus, on a

$$\begin{aligned}
\|v^2 - v^1\| &= \|(1 - \beta) u^2 + \beta \tilde{v}^2 - v^1\| && [\text{par définition de } v^2] \\
&= \|(1 - \beta) u^2 + \beta \tilde{v}^2 + (1 - \beta) (-v^1) + \beta (-v^1)\| \\
&\leq \|(1 - \beta) (u^2 - v^1)\| + \|\beta (\tilde{v}^2 - v^1)\| && [\text{inégalité triangulaire}] \\
&\leq (1 - \beta) \|u^2 - v^1\| + \beta \|\tilde{v}^2 - v^1\| && [0 < \beta \leq 1] \\
&\leq (1 - \beta) (\|u^2\| + \|v^1\|) + \beta \theta \left(\frac{\varepsilon_1}{\rho} + 2 \widetilde{M} \right) \\
&&& [\text{cf. inégalité triangulaire et (1.10)}] \\
&\leq (1 - \beta) \left(\frac{\varepsilon_1}{\rho} + 2 \widetilde{M} \right) + \beta \theta \left(\frac{\varepsilon_1}{\rho} + 2 \widetilde{M} \right)
\end{aligned}$$

$$[\text{cf. } \widetilde{M} \equiv \sup\{\|u\| \mid u \in T(\widetilde{K})\} .$$

D'où $\|u^2\| \leq \widetilde{M}$ (car $u^2 \in T(x^2)$ avec $x^2 \in K \subseteq \widetilde{K}$).

De plus, $\|v^1\| \leq \frac{\varepsilon_1}{\rho} + \widetilde{M}$ en vertu du Lemme 1.2

(en effet, on a bien $v^1 \in T^{\varepsilon_1}(x^1)$ avec $x^1 \in K$)]

$$= \left[\frac{\varepsilon_1}{\rho} + 2 \widetilde{M} \right] [1 - \beta (1 - \theta)] \quad (1.12)$$

Or, par (1.9), on a

$$\frac{1}{\tilde{\varepsilon}_2} \geq \frac{1}{\varepsilon_1 + \frac{\rho \ell}{\rho + \ell} 2 \widetilde{M}}$$

d'où

$$\beta = \frac{\varepsilon_2}{\tilde{\varepsilon}_2} \geq \frac{\varepsilon_2}{\varepsilon_1 + \frac{\rho \ell}{\rho + \ell} 2 \widetilde{M}} .$$

Ainsi,

$$\begin{aligned}
1 - \beta (1 - \theta) &\leq 1 - \frac{\varepsilon_2}{\varepsilon_1 + \frac{\rho \ell}{\rho + \ell} 2 \widetilde{M}} \left(1 - \frac{\ell}{\rho + \ell} \right) & [\theta = \frac{\ell}{\rho + \ell}] \\
&= 1 - \frac{\varepsilon_2}{\varepsilon_1 + \frac{\rho \ell}{\rho + \ell} 2 \widetilde{M}} \frac{\rho}{\rho + \ell} \\
&= 1 - \frac{\varepsilon_2 \rho}{(\rho + \ell) \varepsilon_1 + \rho \ell 2 \widetilde{M}} \\
&= \frac{(\rho + \ell) \varepsilon_1 + \rho \ell 2 \widetilde{M} - \varepsilon_2 \rho}{(\rho + \ell) \varepsilon_1 + \rho \ell 2 \widetilde{M}} \\
&= \frac{\ell [\varepsilon_1 + \rho 2 \widetilde{M}]}{(\rho + \ell) \varepsilon_1 + \rho \ell 2 \widetilde{M}} + \frac{\rho [\varepsilon_1 - \varepsilon_2]}{(\rho + \ell) \varepsilon_1 + \rho \ell 2 \widetilde{M}} \\
&\leq \frac{\ell \varepsilon_1 \rho \left[\frac{1}{\rho} + \frac{2 \widetilde{M}}{\varepsilon_1} \right]}{\rho \varepsilon_1} + \frac{\rho |\varepsilon_1 - \varepsilon_2|}{\rho \varepsilon_1}
\end{aligned}$$

[majoration valable car $\ell, \widetilde{M} \geq 0$ et $\rho, \varepsilon_1 > 0$]

$$\begin{aligned}
&\leq \ell \left[\frac{1}{\rho} + \frac{2 \widetilde{M}}{\varepsilon_1} \right] + \frac{1}{\varepsilon_1} |\varepsilon_1 - \varepsilon_2| \\
&= \|x^1 - x^2\| \left[\frac{1}{\rho} + \frac{2 \widetilde{M}}{\varepsilon_1} \right] + \frac{1}{\varepsilon_1} |\varepsilon_1 - \varepsilon_2| & [\text{cf. } \ell \equiv \|x^1 - x^2\|]
\end{aligned}$$

En remplaçant dans (1.12), on obtient

$$\begin{aligned}
\|v^2 - v^1\| &\leq \left[\frac{\varepsilon_1}{\rho} + 2 \widetilde{M} \right] \left[\|x^1 - x^2\| \left(\frac{1}{\rho} + \frac{2 \widetilde{M}}{\varepsilon_1} \right) + \frac{1}{\varepsilon_1} |\varepsilon_1 - \varepsilon_2| \right] \\
&\leq \left(\frac{\bar{\varepsilon}}{\rho} + 2 \widetilde{M} \right) \left(\frac{1}{\rho} + \frac{2 \widetilde{M}}{\underline{\varepsilon}} \right) \|x^1 - x^2\| + \left(\frac{1}{\rho} + \frac{2 \widetilde{M}}{\underline{\varepsilon}} \right) |\varepsilon_1 - \varepsilon_2| \\
& & [\text{cf. } \varepsilon_1 \in [\underline{\varepsilon}, \bar{\varepsilon}]]
\end{aligned}$$

$$= A \|x^1 - x^2\| + B |\varepsilon_1 - \varepsilon_2| ,$$

c'est-à-dire la thèse. ■

Chapitre 2

Algorithme conceptuel

2.1 Introduction et motivation

Comme annoncé, on va se servir dans ce chapitre de l' ε -élargissement pour décrire un algorithme qui permet de résoudre le problème

$$\text{trouver } x \text{ tel que } 0 \in T(x)$$

où T est un opérateur maximal monotone sur \mathbb{R}^n .

On note l'ensemble des solutions \mathcal{S} et, dans les Chapitres 2 et 3, on suppose $\mathcal{S} \neq \emptyset$.

Cet algorithme est dit **conceptuel** car il utilise les ensembles convexes $T^\varepsilon(x)$. En d'autres mots, on suppose connaître parfaitement $T^\varepsilon(x)$, ce qui est une hypothèse beaucoup trop forte si on veut implémenter l'algorithme (très coûteux, si pas impossible à calculer). Dans le Chapitre 3, on verra une version implémentable de cet algorithme où les ensembles $T^\varepsilon(x)$ seront remplacés par des approximations polyédrales (cf. méthodes faisceaux).

L'idée de base de l'algorithme est assez simple :

Étant donné un point quelconque $y \in \mathbb{R}^n$ et $v \in T(y)$, la monotonie de T implique que

$$\mathcal{S} \subseteq H_{y,v} \stackrel{\text{d\'ef}}{=} \{z \in \mathbb{R}^n \mid \langle z - y, v \rangle \leq 0\}.$$

En effet,

$$z \in \mathcal{S} \quad \Leftrightarrow \quad 0 \in T(z).$$

Comme T est monotone, on a, par définition 1.5,

$$\forall w \in T(z), \forall v \in T(y), \forall y, z \in \mathbb{R}^n : \quad \langle w - v, z - y \rangle \geq 0.$$

Donc, en particulier pour $w = 0$,

$$\begin{aligned}\langle z - y, -v \rangle &\geq 0, \\ \langle z - y, v \rangle &\leq 0.\end{aligned}$$

Notant x^k l'itéré courant (non optimal, c'est-à-dire $0 \notin T(x^k)$), on cherche d'abord y^k et $v^k \in T(y^k)$ tels que

$$x^k \notin H_{y^k, v^k}.$$

Ensuite, on projette x^k sur H_{y^k, v^k} pour obtenir l'itéré suivant, c'est-à-dire

$$x^{k+1} = \text{Proj}_{H_{y^k, v^k}}(x^k) = x^k - \frac{\langle x^k - y^k, v^k \rangle}{\|v^k\|^2} v^k.$$

x^{k+1} étant une projection sur $H_{y^k, v^k} \supseteq \mathcal{S}$, il est plus proche de la solution que x^k .

Autrement dit, on sépare d'abord l'itéré courant des zéros de T en calculant la direction de norme minimum dans $T^{\varepsilon_k}(x^k)$ et l'itéré suivant est donné par la projection de x^k sur l'hyperplan séparant correspondant.

Pour assurer la convergence et une bonne vitesse de convergence (suffisamment de progrès en se déplaçant de x^k vers x^{k+1}), il faut trouver de "bons" y^k et v^k . Comme on suppose que v^k est donné par une sous-routine, le contrôle se fera plutôt sur les y^k . Comme $0 \notin T(x^k)$, on pourrait penser utiliser

$$s^k = \text{Proj}_{T(x^k)}(0)$$

et prendre $y^k = x^k - t_k s^k$ pour un certain $t_k > 0$. Pourtant, parce que $x \mapsto T(x)$ n'est pas continue, ceci est numériquement instable, comme un algorithme de plus forte pente en optimisation non-différentiable, où on peut alors obtenir une meilleure direction s^k en projetant 0 sur un ensemble plus large que $\partial f(x^k)$, à savoir $\partial_\varepsilon f(x^k)$. Dans notre cas général, on va donc procéder par analogie et générer les directions s^k en utilisant $\text{Proj}_{T^{\varepsilon_k}(x^k)}(0)$ pour des valeurs appropriées de ε_k .

Comme alors $y^k = x^k - t_k s^k$, dire qu'il nous faut de "bons" y^k et v^k revient alors à dire qu'il nous faut de "bons" ε_k .

On verra en fait au Théorème 2.1 que $\varepsilon_k \rightarrow 0$.

2.2 L'algorithme conceptuel

Choisir les paramètres $\tau > 0$, $R > 0$, $\varepsilon > 0$, $\sigma \in]0, 1[$.

Poser $k := 0$. Prendre $x^0 \in \mathbb{R}^n$.

Pas # 0 : Test d'arrêt

(0.a) Si $0 \in T(x^k)$, alors STOP.

Pas # 1 : Calcul de la direction de recherche

(1.a) Poser $j := 0$.

(1.b) Calculer $s^{k,j} := \operatorname{argmin} \{ \|v\|^2 \mid v \in T^{\varepsilon 2^{-j}}(x^k) \}$.

(1.c) Si $\|s^{k,j}\| \leq \tau 2^{-j}$, alors : $j := j + 1$,

aller en (1.b).

(1.d) Sinon, définir $j_k := j$, $s^k := s^{k,j_k}$.

Pas # 2 : Recherche linéaire

(2.a) Poser $\ell := 0$.

(2.b) Définir $y^{k,\ell} := x^k - R 2^{-\ell} s^k$ et prendre $v^{k,\ell} \in T(y^{k,\ell})$.

(2.c) Si $\langle v^{k,\ell}, s^k \rangle \leq \sigma \|s^k\|^2$,

alors : $\ell := \ell + 1$,

aller en (2.b).

(2.d) Sinon, définir $\ell_k := \ell$, $v^k := v^{k,\ell_k}$, $y^k := y^{k,\ell_k}$.

Pas # 3 : Projection

(3.a) Définir $x^{k+1} := x^k - \frac{\langle v^k, x^k - y^k \rangle}{\|v^k\|^2} v^k$.

(3.b) Poser $k := k + 1$,

aller en (0.a).

Comme décrit plus haut, le Pas # 1 fournit une direction de recherche telle que

$$\|s^{k,j_k}\| > \tau 2^{-j_k},$$

et le Pas # 2 utilise cette direction pour donner la paire $(y^k, v^k) \in G(T)$ telle que

$$\langle v^k, s^k \rangle > \sigma \|s^k\|^2 ,$$

ce qui assure que $x^k \notin H_{y^k, v^k}$ et aussi que x^{k+1} ne soit pas trop proche de x^k .

Avant de faire une analyse de convergence détaillée, on a besoin de la notion de convergence au sens de Fejer :

Définition 2.1

Une suite $\{x^k\}$ converge vers un ensemble E au sens de Fejer ssi

$$\|x^{k+1} - x\| \leq \|x^k - x\| , \quad \forall x \in E .$$

Proposition 2.1

Soit $\{x^k\}$ une suite qui converge vers un ensemble $E \neq \emptyset$ au sens de Fejer.

Alors,

- 1. la suite $\{x^k\}$ est bornée;*
- 2. si la suite $\{x^k\}$ a un point d'adhérence dans E , alors toute la suite converge vers un point dans E .*

Preuve : voir Théorème 4.1 dans [8]. ■

Remarquons qu'une suite peut être convergente au sens de Fejer sans être convergente au sens de la norme, d'où l'importance du bon choix de y^k et v^k dans notre algorithme pour assurer cette convergence.

2.3 Analyse de convergence

Dans l'algorithme conceptuel, on a trois boucles respectivement sur les indices k , j et ℓ . On va montrer que

- ou bien l'algorithme s'arrête après un nombre fini d'étapes (et alors le dernier point est optimal),
- ou bien l'algorithme génère une suite infinie de points dont aucun n'est optimal (boucle k). On n'aura jamais de boucle infinie pour les indices j et ℓ .

Plus précisément,

Proposition 2.2

Soit x^k l'itéré courant.

Alors, soit x^k est optimal et l'algorithme s'arrête en (0.a) ;

soit x^k n'est pas optimal et on a

1. il existe $j = j_k$ fini tel que

$$\|s^k\| > \tau 2^{-j_k}$$

et j_k est le premier indice à satisfaire cette inégalité, c'est-à-dire

$$\|s^{k,j_k-1}\| \leq \tau 2^{-j_k+1}, \quad \forall j_k > 0.$$

2. il existe $\ell = \ell_k$ fini tel que

$$\langle v^{k,\ell_k}, s^k \rangle > \sigma \|s^k\|^2$$

et ℓ_k est le premier indice à satisfaire cette inégalité, c'est-à-dire

$$\langle v^{k,\ell_k-1}, s^k \rangle \leq \sigma \|s^k\|^2, \quad \forall \ell_k > 0.$$

Preuve :

- Si x^k est optimal, $0 \in T(x^k)$ et l'algorithme s'arrête en (0.a) .
- Si x^k n'est pas optimal,

$$0 \notin T(x^k). \quad (2.1)$$

1. Supposons par l'absurde que l'algorithme boucle à l'infini en j ; alors on aurait une suite infinie $\{s^{k,j}\}_{j \in \mathbb{N}}$ telle que

$$\begin{aligned} s^{k,j} &\in T^{\varepsilon 2^{-j}}(x^k), \quad \text{par} \\ s^{k,j} &= \operatorname{argmin} \{ \|v\|^2 \mid v \in T^{\varepsilon 2^{-j}}(x^k) \} \end{aligned} \quad [\text{cf. (1.b)}]$$

et

$$\|s^{k,j}\| \leq \tau 2^{-j}$$

car on suppose qu'on boucle à l'infini en (1.b)–(1.c) .

En passant à la limite pour $j \rightarrow +\infty$:

$$\|s^{k,j}\| \leq \tau 2^{-j} \rightarrow 0 ,$$

on obtient $\|s^{k,j}\| \rightarrow 0$ et donc $s^{k,j} \rightarrow 0$.

Comme $\varepsilon 2^{-j} \rightarrow 0$, et en vertu de la Proposition 1.5 et de la remarque qui suit, on obtient

$$0 \in T^0(x^k) = T(x^k) ,$$

ce qui est impossible [cf. (2.1)].

D'où il faut qu'il existe j fini tel qu'on sorte de la boucle et, par conséquent, le point 1. est démontré.

2. Supposons par l'absurde que l'algorithme boucle à l'infini en ℓ ; alors on aurait une suite infinie $\{(y^{k,\ell}, v^{k,\ell})\}_{\ell \in \mathbb{N}} \subseteq G(T)$ telle que

$$\begin{aligned} \lim_{\ell \rightarrow +\infty} y^{k,\ell} &= \lim_{\ell \rightarrow +\infty} x^k - R 2^{-\ell} s^k && [\text{cf. (2.b)}] \\ &= x^k && [\text{car } \lim_{\ell \rightarrow +\infty} R 2^{-\ell} s^k = 0] \end{aligned}$$

et

$$\langle v^{k,\ell}, s^k \rangle \leq \sigma \|s^k\|^2 , \quad \forall \ell \in \mathbb{N} \quad (2.2)$$

car on suppose qu'on boucle à l'infini en (2.b)–(2.c) .

$\{y^{k,\ell}\}_{\ell \in \mathbb{N}}$ est donc une suite convergente (vers x^k) et, par conséquent, elle est bornée.

Comme T est maximal monotone, il est localement borné et donc borné sur des parties bornées, en vertu des Propositions 1.7 et 1.6; il s'ensuit que $\{v^{k,\ell}\}_{\ell \in \mathbb{N}} \subseteq T(\{y^{k,\ell}\}_{\ell \in \mathbb{N}})$ est bornée car $\{y^{k,\ell}\}_{\ell \in \mathbb{N}}$ l'est.

On peut donc en extraire une sous-suite $\{v^{k,\ell_i}\}_{i \in \mathbb{N}}$, convergente vers une limite notée \bar{v} .

On a alors

$$\lim_{i \rightarrow +\infty} v^{k,\ell_i} \equiv \bar{v} \in T \left(\lim_{i \rightarrow +\infty} y^{k,\ell_i} \right) = T(x^k)$$

(car $\{y^{k,\ell}\}_{\ell \in \mathbb{N}}$ converge, et donc toute sous-suite converge vers la même limite)
et donc, en vertu de la Proposition 1.4, 1.,

$$\bar{v} \in T(x^k) \subseteq T^{\varepsilon 2^{-j_k}}(x^k) .$$

On a également que, par (2.2) et continuité du produit scalaire,

$$\langle \bar{v}, s^k \rangle \leq \sigma \|s^k\|^2 . \quad (2.3)$$

Mais, comme

$$s^k = \operatorname{argmin} \{ \|v\|^2 \mid v \in T^{\varepsilon 2^{-j}}(x^k) \}$$

c'est-à-dire

$$s^k = \operatorname{Proj}_{T^{\varepsilon 2^{-j}}(x^k)}(0) ,$$

on a, par la caractérisation de la projection P d'un point x sur un ensemble E convexe fermé,

$$\forall y \in E : \quad \langle x - P(x), y - P(x) \rangle \leq 0$$

et donc, pour $x = 0$ et $y = \bar{v} \in T^{\varepsilon 2^{-j}}(x^k)$,

$$\begin{aligned} \langle 0 - s^k, \bar{v} - s^k \rangle &\leq 0 , \\ -\langle s^k, \bar{v} \rangle + \langle s^k, s^k \rangle &\leq 0 , \\ \|s^k\|^2 &\leq \langle \bar{v}, s^k \rangle . \end{aligned} \quad (2.4)$$

Or, pour avoir (2.3) et (2.4), c'est-à-dire

$$\langle \bar{v}, s^k \rangle \leq \sigma \|s^k\|^2 \leq \sigma \langle \bar{v}, s^k \rangle ,$$

il faudrait que $\sigma \geq 1$, ce qui contredit l'hypothèse de l'algorithme selon laquelle $\sigma \in]0, 1[$.

D'où il faut qu'il existe ℓ fini tel qu'on sorte de la boucle et, par conséquent, on a démontré 2. ■

On a maintenant tous les outils pour montrer que l'algorithme se comporte comme décrit plus haut.

Proposition 2.3

Soit x^k l'itéré courant (non optimal).

Alors, après avoir défini x^{k+1} en (3.a), on a

1. $x^k \notin H_{y^k, v^k} \stackrel{\text{déf}}{=} \{z \in \mathbb{R}^n \mid \langle z - y^k, v^k \rangle \leq 0\} ,$
 $x^{k+1} = \operatorname{Proj}_{H_{y^k, v^k}}(x^k) .$
2. $\forall x^* \in \mathcal{S} : \quad \|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 - \|x^{k+1} - x^k\|^2 .$
3. $\|x^{k+1} - x^k\| > R \sigma \tau^2 2^{-\ell_k - 2j_k} \frac{1}{\|v^k\|} .$

Preuve :

1. Comme la Proposition 2.2 montre qu'on n'a jamais de boucle infinie pour les indices j et ℓ , en (3.b), le couple défini comme suit

$$(y^k, v^k) = (x^k - R 2^{-\ell_k} s^k, v^{k, \ell_k})$$

est tel que

$$\begin{aligned} \langle x^k - y^k, v^k \rangle &= \langle x^k - x^k + R 2^{-\ell_k} s^k, v^{k, \ell_k} \rangle \\ &= R 2^{-\ell_k} \langle s^k, v^{k, \ell_k} \rangle \\ &> R 2^{-\ell_k} \sigma \|s^k\|^2 \quad [\text{car on est sorti de la boucle en (2.c)}] \\ &> R 2^{-\ell_k} \sigma (\tau 2^{-j_k})^2 \quad [\text{car on est sorti de la boucle en (1.c)}] \\ &= R \sigma \tau^2 2^{-\ell_k - 2j_k} \\ &> 0, \end{aligned}$$

d'où

$$x^k \notin H_{y^k, v^k},$$

et on obtient

$$x^{k+1} = \text{Proj}_{H_{y^k, v^k}}(x^k)$$

car, par définition en (3.a),

$$x^{k+1} = x^k - \frac{\langle v^k, x^k - y^k \rangle v^k}{\|v^k\|^2}.$$

2. La caractérisation de la projection P d'un point x sur un ensemble E convexe fermé

$$\forall y \in E : \quad \langle x - P(x), y - P(x) \rangle \leq 0$$

donne

$$\forall x^* \in \mathcal{S} \subseteq H_{y^k, v^k} : \quad \langle x^k - x^{k+1}, x^* - x^{k+1} \rangle \leq 0$$

et donc

$$\begin{aligned} \|x^k - x^*\|^2 &= \|(x^k - x^{k+1}) - (x^* - x^{k+1})\|^2 \\ &= \|x^k - x^{k+1}\|^2 + \|x^* - x^{k+1}\|^2 - 2 \underbrace{\langle x^k - x^{k+1}, x^* - x^{k+1} \rangle}_{\leq 0} \\ &\geq \|x^k - x^{k+1}\|^2 + \|x^* - x^{k+1}\|^2 \\ \|x^{k+1} - x^*\|^2 &\leq \|x^k - x^*\|^2 - \|x^{k+1} - x^k\|^2. \end{aligned}$$

3. Par définition de x^{k+1} ,

$$\begin{aligned}\|x^{k+1} - x^k\| &= \left\| -\frac{\langle v^k, x^k - y^k \rangle v^k}{\|v^k\|^2} \right\| \\ &= \frac{|\langle v^k, x^k - y^k \rangle| \|v^k\|}{\|v^k\|^2}.\end{aligned}$$

En reprenant la même minoration qu'au point 1., on a immédiatement

$$\|x^{k+1} - x^k\| > R \sigma \tau^2 2^{-\ell_k - 2j_k} \frac{1}{\|v^k\|} . \quad \blacksquare$$

Il est maintenant assez facile de prouver la convergence au sens de Fejer ainsi que le caractère borné des variables de l'algorithme.

Lemme 2.1

Si l'algorithme conceptuel génère une suite infinie $\{x^k\}$, alors celle-ci converge au sens de Fejer vers l'ensemble des solutions \mathcal{S} .

De plus, x^k , s^k , y^k et v^k , de même que les ensembles $\{s^{k,j}\}_{0 \leq j \leq j_k}$ et $\{(y^{k,\ell}, v^{k,\ell})\}_{0 \leq \ell \leq \ell_k}$ sont bornés.

Preuve : Remarquons que, par définition,

$$s^k = s^{k,j_k} , \quad [\text{cf. (1.d)}]$$

$$y^k = y^{k,\ell_k} \quad \text{et} \quad v^k = v^{k,\ell_k} . \quad [\text{cf. (2.d)}]$$

Il suffit donc de montrer que $\{s^{k,j}\}_{0 \leq j \leq j_k}$ et $\{(y^{k,\ell}, v^{k,\ell})\}_{0 \leq \ell \leq \ell_k}$ sont bornés car alors, s^k , y^k et v^k le seront aussi puisque ce sont des éléments de ces ensembles.

Par la Proposition 2.3, 2., on obtient

$$\begin{aligned}\forall x^* \in \mathcal{S} : \quad \|x^{k+1} - x^*\|^2 &\leq \|x^k - x^*\|^2 - \|x^{k+1} - x^k\|^2 \\ &\leq \|x^k - x^*\|^2 \quad [\text{car } \|x^{k+1} - x^k\|^2 \geq 0];\end{aligned}$$

d'où, par définition, $\{x^k\}$ converge au sens de Fejer vers l'ensemble \mathcal{S} .

En vertu de la Proposition 2.1, 1., la suite $\{x^k\}$ est alors également bornée, c'est-à-dire

$$\exists c \text{ tel que } \|x_k\| \leq c , \quad \forall k .$$

Il existe donc un compact K_0 (prendre par exemple $K_0 = \overline{B(0, c)}$) tel que

$$\{x^k\} \subseteq K_0. \quad (2.5)$$

Comme T est maximal monotone, il est localement borné en vertu de la Proposition 1.7. Par la remarque qui suit le Théorème 1.2, T^ε est alors localement borné et donc borné sur des parties bornées par la Proposition 1.6.

Ainsi $T^\varepsilon(K_0)$ est borné et $M_0 \equiv \sup\{\|u\| \mid u \in T^\varepsilon(K_0)\}$ est fini.

Par conséquent, $s^{k,j}$ est borné pour tout $j \leq j_k$ c'est-à-dire

$$\|s^{k,j}\| \leq M_0, \quad \forall 0 \leq j \leq j_k;$$

ou encore

$$\{s^{k,j}\} \subseteq B(0, M_0), \quad \forall 0 \leq j \leq j_k; \quad (2.6)$$

en effet,

$$s^{k,j} = \operatorname{argmin} \{\|v\|^2 \mid v \in T^{\varepsilon \cdot 2^{-j}}(x^k)\}$$

et on a donc bien, par la Proposition 1.4, 1.,

$$s^{k,j} \in T^{\varepsilon \cdot 2^{-j}}(x^k) \subseteq T^\varepsilon(x^k).$$

En ce qui concerne $\{(y^{k,\ell}, v^{k,\ell})\}_{0 \leq \ell \leq \ell_k}$, on a, par définition (2.b),

$$y^{k,\ell} = x^k - 2^{-\ell} R s^k.$$

Comme $\{x^k\} \subseteq K_0$ [cf. (2.5)], $\ell_k \geq 0$ et $\{s^k\} \subseteq B(0, M_0)$ [cf. (2.6)], on a, en définissant $K_1 \equiv K_0 + \overline{B(0, R M_0)}$,

$$y^{k,\ell} \in K_1, \quad \forall 0 \leq \ell \leq \ell_k.$$

Finalement, comme T est localement borné et donc borné sur des parties bornées, il s'ensuit que $T(\{y^{k,\ell}\}_{0 \leq \ell \leq \ell_k}) \supseteq \{v^{k,\ell}\}_{0 \leq \ell \leq \ell_k}$ est borné. ■

Avant de prouver la convergence de l'algorithme, on prouve encore un lemme technique préliminaire :

Lemme 2.2

Supposons que l'algorithme conceptuel génère une suite infinie $\{x^k\}$.

Si la suite $\{j_k\}_{k \in \mathbb{N}}$ est bornée, alors la suite $\{\ell_k\}_{k \in \mathbb{N}}$ est également bornée.

Preuve : Posons $\varepsilon_k \equiv \varepsilon 2^{-j_k}$.

$\{j_k\}_{k \in \mathbb{N}}$ est bornée par hypothèse; d'où

$$\exists J \text{ tel que } j_k \leq J, \quad \forall k \in \mathbb{N}$$

et

$$\varepsilon_k \in [\varepsilon 2^{-J}, \varepsilon].$$

Pour les indices k tels que $\ell_k > 0$, les couples $(y^{k, \ell_k-1}, v^{k, \ell_k-1})$ [bien définis car $\ell_k > 0 \Rightarrow \ell_k - 1 \geq 0$] satisfont

$$\begin{aligned} \|y^{k, \ell_k-1} - x^k\| &= \|x^k - R 2^{-(\ell_k-1)} s^k - x^k\| & [\text{cf. (2.b)}] \\ &= R 2^{-\ell_k+1} \|s^k\| & (2.7) \end{aligned}$$

et

$$\begin{aligned} v^{k, \ell_k-1} &\in T(y^{k, \ell_k-1}) & [\text{cf. (2.b)}] \\ &\subseteq T^{\varepsilon_k}(y^{k, \ell_k-1}). & (2.8) \end{aligned}$$

[cf. Proposition 1.4, 1.]

Appliquons le Théorème 1.5 à

$$x^1 = y^{k, \ell_k-1} \quad \text{borné en vertu du Lemme 2.1,}$$

$$x^2 = x^k \quad \text{borné en vertu du Lemme 2.1,}$$

$$\left. \begin{array}{l} \varepsilon_1 = \varepsilon_k, \\ \varepsilon_2 = \varepsilon_k, \end{array} \right\} \quad \varepsilon_k \in [\varepsilon 2^{-J}, \varepsilon] = [\underline{\varepsilon}, \bar{\varepsilon}],$$

$$v^1 = v^{k, \ell_k-1}, \quad v^1 \in T^{\varepsilon_1}(x^1) \quad \text{par (2.8).}$$

Par le Théorème,

$$\exists v^2 \in T^{\varepsilon_2}(x^2) = T^{\varepsilon_k}(x^k) \quad (2.9)$$

tel que, pour deux constantes positives A et B ,

$$\|v^1 - v^2\| \leq A \|x^1 - x^2\| + B |\varepsilon_1 - \varepsilon_2|,$$

ce qui donne, ensemble avec (2.7)

$$\begin{aligned} \|v^{k, \ell_k-1} - v^2\| &\leq A \|y^{k, \ell_k-1} - x^k\| + B |\varepsilon_k - \varepsilon_k| \\ &\leq A R 2^{-\ell_k+1} \|s^k\| \quad \text{pour chaque } k \text{ tel que } \ell_k > 0. \end{aligned} \quad (2.10)$$

Considérons ensuite la projection

$$\bar{v}^k \stackrel{\text{déf}}{=} \text{Proj}_{T^{\varepsilon_k}(x^k)} (v^{k,\ell_k-1}) .$$

Il est clair que

$$\|s^k\| \leq \|\bar{v}^k\| \quad (2.11)$$

car

$$\begin{aligned} s^k &= s^{k,j_k} \\ &= \text{argmin} \{ \|v\|^2 \mid v \in T^{\varepsilon_{2^{-j_k}}}(x^k) = T^{\varepsilon_k}(x^k) \} \end{aligned}$$

et, par définition,

$$\bar{v}^k \in T^{\varepsilon_k}(x^k) .$$

De plus, la caractérisation de la projection P d'un point x sur un ensemble E convexe fermé

$$\forall y \in E : \quad \langle x - P(x), y - P(x) \rangle \leq 0$$

donne

$$\forall w \in T^{\varepsilon_k}(x^k) : \quad \langle v^{k,\ell_k-1} - \bar{v}^k, w - \bar{v}^k \rangle \leq 0$$

et donc

$$\begin{aligned} \|v^{k,\ell_k-1} - w\|^2 &= \|-w + v^{k,\ell_k-1} - \bar{v}^k + \bar{v}^k\|^2 \\ &= \|-w + \bar{v}^k\|^2 + \|v^{k,\ell_k-1} - \bar{v}^k\|^2 - 2 \underbrace{\langle w - \bar{v}^k, v^{k,\ell_k-1} - \bar{v}^k \rangle}_{\leq 0} \\ &\geq \|v^{k,\ell_k-1} - \bar{v}^k\|^2 + \|-w + \bar{v}^k\|^2 \\ &\geq \|v^{k,\ell_k-1} - \bar{v}^k\|^2, \quad \forall w \in T^{\varepsilon_k}(x^k) . \end{aligned}$$

En particulier, pour $w = v^2 \in T^{\varepsilon_k}(x^k)$ [cf. (2.9)],

$$\begin{aligned} \|v^{k,\ell_k-1} - \bar{v}^k\| &\leq \|v^{k,\ell_k-1} - v^2\| \\ &\leq A R 2^{-\ell_k+1} \|s^k\| && \text{[cf. (2.10)]} \\ &\leq A R 2^{-\ell_k+1} \|\bar{v}^k\| && \text{[cf. (2.11)]} \\ &\leq A R 2^{-\ell_k+1} \|\bar{v}^k\| \quad \text{pour chaque } k \text{ tel que } \ell_k > 0 . && (2.12) \end{aligned}$$

Considérons

$$\begin{aligned}
 \langle v^{k, \ell_k - 1}, s^k \rangle &= \langle v^{k, \ell_k - 1} - \bar{v}^k, s^k \rangle + \langle \bar{v}^k, s^k \rangle \\
 &\geq -\|v^{k, \ell_k - 1} - \bar{v}^k\| \|s^k\| + \langle \bar{v}^k, s^k \rangle && \text{[Cauchy-Schwarz]} \\
 &\geq -A R 2^{-\ell_k + 1} \|\bar{v}^k\| \|s^k\| + \langle \bar{v}^k, s^k \rangle . && \text{[cf. (2.12)]} \\
 &\geq -A R 2^{-\ell_k + 1} \|s^k\|^2 + \langle \bar{v}^k, s^k \rangle . && \text{[cf. (2.11)]}
 \end{aligned}$$

Or, la caractérisation de la projection donne, pour $x = 0$, $P(x) = s^k$ et $y = \bar{v}^k$:

$$\begin{aligned}
 \langle 0 - s^k, \bar{v}^k - s^k \rangle &\leq 0 \\
 \|s^k\|^2 &\leq \langle \bar{v}^k, s^k \rangle ,
 \end{aligned}$$

ce qui permet d'écrire

$$\langle v^{k, \ell_k - 1}, s^k \rangle \geq [1 - A R 2^{-\ell_k + 1}] \|s^k\|^2 .$$

Par la Proposition 2.2, 2., on sait que

$$\langle v^{k, \ell_k - 1}, s^k \rangle \leq \sigma \|s^k\|^2 .$$

Par conséquent,

$$\begin{aligned}
 \sigma &\geq 1 - A R 2^{-\ell_k + 1} , \\
 -A R 2^{-\ell_k + 1} &\leq \sigma - 1 , \\
 2^{-\ell_k + 1} &\geq \frac{1 - \sigma}{A R} , \\
 -\ell_k + 1 &\geq \log_2 \frac{1 - \sigma}{A R} , \\
 -\ell_k &\geq \log_2 \frac{1 - \sigma}{A R} - \log_2 2 , \\
 \ell_k &\leq -\log_2 \frac{1 - \sigma}{2 A R} , \\
 \ell_k &\leq \log_2 \frac{2 A R}{1 - \sigma} \quad \text{pour chaque } k \text{ tel que } \ell_k > 0 ,
 \end{aligned}$$

c'est-à-dire

$$\ell_k \leq \max \left\{ 0, \log_2 \frac{2 A R}{1 - \sigma} \right\} .$$

Par conséquent, on obtient bien que la suite $\{\ell_k\}_{k \in \mathbb{N}}$ est bornée. ■

Théorème 2.1

Considérons la suite $\{x^k\}$ générée par l'algorithme conceptuel.

Alors

- ou bien, elle est finie, et le dernier élément est optimal;*
- ou bien elle est infinie, et elle converge vers une solution du problème :*

trouver x tel que $0 \in T(x)$.

Preuve :

- Le cas fini a été traité dans la Proposition 2.2.
- Si la suite est infinie, alors aucun x^k n'est optimal et, par la Proposition 2.3, 2., on a

$$\forall x^* \in \mathcal{S} : \quad \|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 - \|x^{k+1} - x^k\|^2 ,$$

c'est-à-dire

$$\begin{aligned} \|x^1 - x^0\|^2 &\leq \|x^0 - x^*\|^2 - \|x^1 - x^0\|^2 , \\ \|x^2 - x^1\|^2 &\leq \|x^1 - x^*\|^2 - \|x^2 - x^1\|^2 , \\ \|x^3 - x^2\|^2 &\leq \|x^2 - x^*\|^2 - \|x^3 - x^2\|^2 , \\ &\vdots \\ \|x^{n+1} - x^n\|^2 &\leq \|x^n - x^*\|^2 - \|x^{n+1} - x^n\|^2 . \end{aligned}$$

En sommant,

$$\begin{aligned} \sum_{k=0}^n \|x^{k+1} - x^k\|^2 &\leq \|x^0 - x^*\|^2 - \|x^{n+1} - x^*\|^2 \\ &\leq \|x^0 - x^*\|^2 . \end{aligned} \quad [\text{car } \|x^{n+1} - x^*\|^2 \geq 0]$$

En faisant tendre $n \rightarrow +\infty$, on obtient

$$\sum_{k=0}^{+\infty} \|x^{k+1} - x^k\|^2 < +\infty$$

car le membre de droite est un nombre fini indépendant de n .

La condition nécessaire de convergence nous permet alors d'affirmer que

$$\lim_{k \rightarrow +\infty} \|x^{k+1} - x^k\| = 0 . \quad (2.13)$$

Par la Proposition 2.3, 3., on sait que

$$\begin{aligned} \|x^{k+1} - x^k\| &> R \sigma \tau^2 2^{-\ell_k - 2j_k} \frac{1}{\|v^k\|} \\ &> \frac{R \sigma \tau^2}{M} 2^{-\ell_k - 2j_k} \end{aligned}$$

[cf. v^k est borné en vertu du Lemme 2.1, et on note M une borne supérieure].

En laissant tendre $k \rightarrow +\infty$, on obtient, par (2.13),

$$0 \geq \lim_{k \rightarrow +\infty} \frac{R \sigma \tau^2}{M} 2^{-\ell_k - 2j_k} \geq 0$$

et donc

$$0 \geq \lim_{k \rightarrow +\infty} 2^{-\ell_k - 2j_k} \geq 0 .$$

[en effet, $R > 0$, $\tau > 0$, $\sigma \in]0, 1[$ par hypothèse,

de plus, $M > 0$ car $\|v^k\| \leq M$, $\forall k$

et on ne peut pas avoir $v^k = 0$ car, pour sortir des boucles en ℓ et j (dont on a montré qu'elles étaient toujours finies), il faut que

$$\langle v^k, s^k \rangle > \sigma \|s^k\|^2 \quad (2.c) \quad \text{et} \quad \|s^k\| > \tau 2^{-j} > 0 \quad (1.c)]$$

Ainsi, en vertu du Théorème de l'Étau,

$$\lim_{k \rightarrow +\infty} 2^{-\ell_k - 2j_k} = 0$$

et, par conséquent,

$$\lim_{k \rightarrow +\infty} \ell_k + 2 j_k = +\infty . \quad (2.14)$$

Or, d'après le Lemme 2.2: si $\{j_k\}_{k \in \mathbb{N}}$ est bornée, alors $\{\ell_k\}_{k \in \mathbb{N}}$ est également bornée, d'où, comme on a (2.14), $\{j_k\}_{k \in \mathbb{N}}$ ne peut pas être bornée.

Comme $\{x^k\}$ est une suite infinie, par le Lemme 2.1, elle converge au sens de Fejer, et, par la Proposition 2.1, 1., elle est alors bornée et on peut donc en extraire une sous-suite convergente (au sens de la norme):

$$\lim_{q \rightarrow +\infty} x^{k_q} \stackrel{\text{not}}{=} \hat{x} .$$

Comme $\{j_k\}_{k \in \mathbb{N}}$ n'est pas bornée, c'est-à-dire

$$\lim_{q \rightarrow +\infty} j_{k_q} = +\infty.$$

$$\lim_{k \rightarrow +\infty} j_k = +\infty ,$$

on obtient également le résultat suivant :

$$\lim_{q \rightarrow +\infty} j_{k_q} = +\infty , \quad j_{k_q} > 0 , \quad \forall q .$$

En vertu de la Proposition 2.2, 1.,

$$0 \leq \|s^{k_q, j_{k_q}-1}\| \leq \tau 2^{-j_{k_q}+1} ;$$

en laissant tendre $q \rightarrow +\infty$,

$$0 \leq \lim_{q \rightarrow +\infty} \|s^{k_q, j_{k_q}-1}\| \leq 0$$

par conséquent,

$$\lim_{q \rightarrow +\infty} \|s^{k_q, j_{k_q}-1}\| = 0$$

et donc

$$\lim_{q \rightarrow +\infty} s^{k_q, j_{k_q}-1} = 0 .$$

Comme, de plus,

$$\lim_{q \rightarrow +\infty} x^{k_q} = \hat{x} ,$$

$$\lim_{q \rightarrow +\infty} \varepsilon 2^{-j_{k_q}+1} = 0 \quad [\text{car } \lim_{q \rightarrow +\infty} j_{k_q} = +\infty]$$

et qu'on sait, par définition que

$$s^{k_q, j_{k_q}-1} \in T^{\varepsilon 2^{-j_{k_q}+1}}(x^{k_q}) , \quad [\text{cf. (1.b)}]$$

on obtient, en vertu de la Proposition 1.5,

$$0 \in T^0(\hat{x}) = T(\hat{x}) \quad \text{c'est-à-dire } \hat{x} \in \mathcal{S} .$$

La suite $\{x^k\}$ convergente au sens de Fejer a donc un point d'adhérence appartenant à \mathcal{S} ,
et donc, par le Proposition 2.1, 2., toute la suite converge vers un point dans \mathcal{S} . \blacksquare

Chapitre 3

Algorithme implémentable

3.1 Introduction et motivation

Comme annoncé au Chapitre 2, on va maintenant aborder la version implémentable de l'algorithme en vue de résoudre le problème suivant :

$$\text{trouver } x \text{ tel que } 0 \in T(x)$$

où T est un opérateur maximal monotone sur \mathbb{R}^n . Au Chapitre 2, on a surtout porté l'attention sur le fait d'avoir de "bons" ε_k ($\varepsilon_k \rightarrow 0$) pour assurer la convergence. Pratiquement, il se fait qu'il vaut mieux ne pas laisser tendre ε_k vers 0 trop rapidement, car les T^{ε_k} correspondants pourraient être insuffisamment réguliers et alors, x^{k+1} serait trop proche de x^k (on ne progresse presque pas).

Un autre problème déjà mentionné est qu'on a défini

$$s_k = \text{Proj}_{T^{\varepsilon_k}(x^k)}(0),$$

ce qui suppose qu'on connaît parfaitement $T^{\varepsilon_k}(x^k)$, ce qui est une hypothèse beaucoup trop forte si on veut implémenter l'algorithme. C'est ici qu'entre en jeu la méthode faisceau :

À chaque itération, on a un faisceau avec toutes les informations qu'on possède à ce moment-là

$$\{(z^i, w^i \in T(z^i))\}_{i \leq p}.$$

L'enveloppe convexe de quelques w^i bien choisis est une approximation de $T^\varepsilon(x^k)$. La direction de recherche s^k peut alors être approchée par la projection de 0 sur cette approximation polyédrale de $T^\varepsilon(x^k)$.

Il faut donc bien choisir le sous-faisceau pour avoir de “bons” ε_k et assurer la convergence.

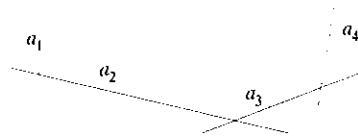
Lorsque $T \neq \partial f$, on ne dispose pas de valeurs fonctionnelles. Il faudra donc adapter la méthode faisceau de l’optimisation à ce cas plus général.

Avant d’aborder ce problème, rappelons d’abord l’idée de la méthode faisceau pour résoudre $0 \in \partial f(x)$ et indiquons les généralisations possibles :

En chaque point $x \in \mathbb{R}^n$, la seule information dont on dispose est la valeur de la fonction objectif en ce point et un sous-gradient de f en x .

À la k -ième itération, le faisceau de sous-gradients $\{g^1, \dots, g^k\}$ calculé aux itérés x^1, \dots, x^k respectivement, est utilisé pour approcher le sous-différentiel de f en x^k , et la nouvelle direction de descente est calculée comme l’opposé de la projection de 0 sur cette approximation. Si celle-ci est bonne, on effectue une recherche linéaire suivant la direction, sinon on l’améliore en ajoutant au faisceau $\{g^1, \dots, g^k\}$ un nouveau sous-gradient de f , calculé en un point voisin de x^k . Ceci évite de devoir calculer tout le sous-différentiel, ce qui est très coûteux (si pas impossible), sauf si f a une forme très particulière. En effet, si, par exemple f est le maximum d’un nombre fini de fonctions linéaires :

$$f(x) = \max\{a_i^T x + b_i \mid 1 \leq i \leq k\},$$



alors

$$\partial f(x) = \text{co} \{a_j \mid f(x) = a_j^T x + b_j\}.$$

Comme, dans ce cas, on dispose d’une valeur fonctionnelle, l’idée pour résoudre le problème

$$\text{trouver } x \text{ tel que } 0 \in \partial f(x)$$

est de trouver une suite $\{x^k\}$ telle que $\{f(x^k)\}$ décroît et telle que au moins un point d’adhérence \bar{x} est tel que $0 \in \partial f(\bar{x})$. Ce sera le cas si $\{x^k\}$ est bornée et 0 est un point d’adhérence de la suite $\{D^k\}$ des distances euclidiennes entre 0 et l’ensemble convexe fermé $\partial f(x^k)$. En effet, si $g^k \in \partial f(y^k)$ pour tout k et si $g^k \rightarrow \bar{g}$ et $y^k \rightarrow \bar{y}$, alors $\bar{g} \in \partial f(\bar{y})$.

Ceci peut se faire en prenant

$$x^{k+1} = x^k + t_k d^k$$

où

$$d^k \equiv - \operatorname{argmin} \{ \|v\| \mid v \in \partial f(x^k) \}$$

est une direction de descente si $0 \notin \partial f(x^k)$ et t_k une longueur de pas.

Malheureusement, il se peut qu'il n'existe pas de \bar{x} tel que $0 \in \partial f(\bar{x})$ et que 0 ne soit pas point d'adhérence de $\{D^k\}$. Ceci est dû au fait que, même si f est convexe, $\partial f(x)$ n'est pas nécessairement semi-continue inférieurement, c'est-à-dire on n'a **pas** que, si $y^k \rightarrow \bar{y}$ et $\bar{g} \in \partial f(\bar{y})$, alors il existe $\{g^k\}$ tel que $g^k \in \partial f(y^k)$ pour tout k . $g^k \rightarrow \bar{g}$

Cependant, si f est convexe, il existe une approximation du sous-différentiel qui est bornée et semi-continue inférieurement : l' ε -sous-différentiel [cf. Définition 1.9].

▽ Remarquons que, dans notre cas général, l' ε -élargissement T^ε est Lipschitz continu [cf. Théorème 1.4], propriété plus forte que la semi-continuité inférieure.

△

L'idée est alors de générer deux suites $\{x^k\}$ et $\{\varepsilon_k\}$ telles que $\{f(x^k)\}$ décroît, $\varepsilon_k \rightarrow 0$ et 0 est point d'adhérence de la suite $\{E^k\}$ des distances euclidiennes entre 0 et $\partial_{\varepsilon_k} f(x^k)$.

Si $0 \in \partial_{\varepsilon_k} f(x^k)$, x^k est solution à ε_k près, c'est-à-dire

$$f(x^k) \leq f(x) + \varepsilon_k, \quad \forall x \in \mathbb{R}^n.$$

Sinon,

$$d^k = - \operatorname{argmin} \{ \|v\| \mid v \in \partial_{\varepsilon_k} f(x^k) \}$$

est une direction de descente.

Un algorithme simple est alors de garder ε constant jusqu'à avoir une solution à ε près, puis de diminuer ε :

1. $d^k = - \operatorname{argmin} \{ \|v\| \mid v \in \partial_{\varepsilon_k} f(x^k) \};$

2. si $d^k = 0$, alors, $\varepsilon_{k+1} = \frac{\varepsilon_k}{2}$,

- sinon $\varepsilon_{k+1} = \varepsilon_k$.

3. $x^{k+1} = x^k + t_k d^k$

où on effectue une recherche linéaire pour trouver t_k qui donne une décroissance suffisante de f .

▽ Ceci est proche de notre algorithme conceptuel :

$$1. s^k = \operatorname{argmin} \{ \|v\|^2 \mid v \in T^{\varepsilon_j}(x^k) \} .$$

$$2. \text{ si } \|s^k\| \leq \tau 2^{-j} , \text{ alors } \varepsilon_{j+1} = \frac{\varepsilon_j}{2} ;$$

sinon, effectuer une recherche linéaire

$$y^{k,\ell} = x^k - t_\ell s^k , \quad v^{k,\ell} \in T(y^{k,\ell})$$

pour trouver t_ℓ qui assure que $\langle v^{k,\ell}, s^k \rangle > \sigma \|s^k\|^2$ pour que x^{k+1} ne soit pas trop proche de x^k .

$$3. x^{k+1} = \operatorname{Proj}_{H_{y^k, v^k}}(x^k) .$$

△

Le problème est qu'on suppose connaître parfaitement $\partial_{\varepsilon_k} f(x^k)$, ce qui n'est pas le cas. Pour avoir un algorithme implémentable, il faut donc le remplacer par un autre ensemble convexe qu'on peut traiter numériquement.

▽ Dans l'algorithme conceptuel du Chapitre 2, on a aussi supposé connaître parfaitement $T^\varepsilon(x^k)$. Le but de ce Chapitre 3 est précisément de le remplacer par un autre ensemble qu'on peut traiter numériquement pour avoir un algorithme implémentable.

△

Dans la méthode faisceau, on suppose qu'en chaque point x^i , la valeur de la fonction $f(x^i)$ et un sous-gradient $g^i \in \partial f(x^i)$ peuvent être évalués.

Ainsi, pour l'approximation, on dispose du faisceau $\{(x^i, f(x^i), g^i \in \partial f(x^i))\}_{i \leq k}$.

▽ Dans le cas des opérateurs, on supposera donc qu'en chaque point z , un $w \in T(z)$ peut être évalué.

Ainsi, pour l'approximation, on dispose du faisceau $\{(z^i, w^i \in T(z^i))\}_{i \leq p}$.

△

On approche alors f par l'approximation polyédrale convexe suivante :

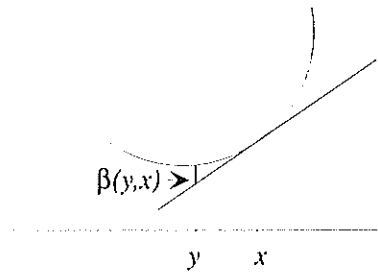
$$\begin{aligned} \bar{f}_k(x) &= \max_{1 \leq i \leq k} \{ f(x^i) + \langle g^i, x - x^i \rangle \} \\ (&= \max_{1 \leq i \leq k} \{ f(x^k) - \beta(x^k, x^i) + \langle g^i, x - x^i \rangle \}) \end{aligned}$$

où

$$\beta(y, x) \stackrel{\text{def}}{=} f(y) - \{f(x) + \langle g(x), y - x \rangle\}$$

(avec $g(x)$ sous-gradient de $\partial f(x)$ qui est disponible)

peut être interprété comme une mesure de combien le sous-gradient $g(x) \in \partial f(x)$ diffère d'être un élément de $\partial f(y)$. En d'autres termes, $\beta(y, x)$ est l'erreur de linéarisation faite en y lorsque f est linéarisée en x avec une pente $g(x)$.



On a ainsi $g(x) \in \partial_{\beta(y, x)} f(y)$.

À cause de la forme très particulière de \bar{f}_k , il est alors facile de donner l' ε -sous-différentiel (cf. page 53)

$$\partial_\varepsilon \bar{f}_k(x^k) = \left\{ \sum_{i=1}^k \lambda_i g^i \mid \lambda_i \geq 0, \sum_{i=1}^k \lambda_i = 1, \sum_{i=1}^k \lambda_i \beta(x^k, x^i) \leq \varepsilon \right\}.$$

On définit alors la direction le long de laquelle on effectue la recherche linéaire par

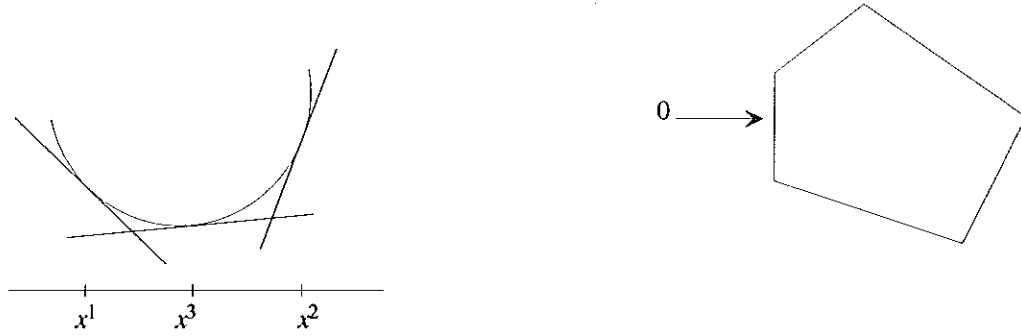
$$d^k = - \operatorname{argmin} \{ \|v\| \mid v \in \partial_\varepsilon \bar{f}_k(x^k) \};$$

on l'obtient en résolvant le problème

$$\left\{ \begin{array}{l} \min \frac{1}{2} \left\| \sum_{i=1}^k \lambda_i g^i \right\|^2 \\ \text{sous les contraintes} \quad \sum_{i=1}^k \lambda_i = 1, \\ \lambda_i \geq 0, \quad i = 1, \dots, k, \\ \sum_{i=1}^k \lambda_i \beta(x^k, x^i) \leq \varepsilon; \end{array} \right.$$

alors,

$$d^k = - \sum_{i=1}^k \lambda_i^* g^i.$$



▽ Dans le cas général, $T^{\varepsilon 2^{-j}}(x^k)$ sera approché par une combinaison convexe d'éléments choisis parmi $\{(z^i, w^i \in T(z^i))\}_{i \leq p}$. En fait, on retient les éléments pour lesquels les antécédents ne sont pas trop éloignés de x^k . C'est le sous-faisceau défini par les indices

$$I_{k,n,j} = \{1 \leq i \leq p \mid \|z^i - x^k\| \leq R 2^{-j}\}$$

(la signification de n apparaîtra plus tard).

On définit alors $s^{k,n}$ comme la projection de 0 sur cette approximation polyédrale de $T^{\varepsilon 2^{-j}}(x^k)$ qui, comme dans l'algorithme conceptuel, satisfait $\|s^{k,n}\| > \tau 2^{-j}$; on l'obtient en résolvant

$$\begin{cases} \min \left\| \sum_{i \in I_{k,n,j}} \alpha_i w^i \right\|^2 \\ \text{sous les contraintes } \sum_{i \in I_{k,n,j}} \alpha_i = 1 ; \end{cases}$$

alors,

$$s^{k,n,j} = \sum_{i \in I_{k,n,j}} \alpha_i^* w^i .$$

Si

$$\|s^{k,n,j}\| \leq \tau 2^{-j} ,$$

on fait $j := j + 1$, ce qui définit un nouvel ensemble $I_{k,n,j}$ et on recommence jusqu'à ce qu'on trouve $s^{k,n,j_k,n} = s^{k,n}$ tel que $\|s^{k,n}\| > \tau 2^{-j}$.

△

Si $0 \notin \partial_{\varepsilon} \bar{f}_k(x^k)$, alors d^k est une direction de ε -descente pour l'approximation \bar{f}_k mais non nécessairement pour f . Si ce n'est pas le cas, il faut améliorer l'approximation.

On procède d'après la règle suivante:

SI la direction de descente permet une décroissance suffisante de f , et

SI le pas n'est pas trop court,
autrement dit,

SI il est possible de trouver t tel que

$$f(x^k + t d^k) \leq f(x^k) + m_1 t v^k$$

et

$$\langle g(x^k + t d^k), d^k \rangle \geq m_2 v^k$$

avec $0 < m_1 < m_2 < 1$ et $v^k = -\|d^k\|^2 - \mu_k \varepsilon < 0$,

ALORS on obtient un nouveau point $x^{k+1} = x^k + t_k d^k$ et

on parle de PAS SÉRIEUX.

SINON, on ne bouge pas, c'est-à-dire $x^{k+1} = x^k$ et un nouveau sous-gradient $g(x^k + t d^k)$

[choisi tel que $\beta(x^k, x^k + t d^k) \leq m_3 \varepsilon$ avec $0 < m_3 < 1$]

est ajouté au faisceau $\{g^1, \dots, g^k\}$ tel que l'approximation $\partial_\varepsilon \bar{f}_{k+1}(x^{k+1})$ est une meilleure approximation de $\partial f(x^{k+1})$.

On parle alors de PAS NUL.

∇ Pour les opérateurs, $s^{k,n}$ n'est plus nécessairement l'élément de norme minimum dans $T^{\varepsilon 2^{-j}}(x^k)$. On ne peut donc plus reprendre la preuve de la Proposition 2.2, 2., et on doit ajouter un autre test d'arrêt pour la boucle en j .

On procède comme suit :

SI $(\langle v^{k,n,\ell}, s^{k,n} \rangle \leq \sigma \|s^{k,n}\|^2 \text{ ET } \ell < j_{k,n} + 1)$;

ALORS $\ell = \ell + 1$,

SINON, on sort de la boucle.

SI on sort de la boucle parce que $\ell \geq j_{k,n} + 1$ et $\langle v^{k,n}, s^{k,n} \rangle \leq \sigma \|s^{k,n}\|^2$,

ALORS on n'est plus sûr que x^{k+1} n'est pas trop proche de x^k .

Dans ce cas, il faut améliorer l'approximation, ce qui sera fait d'après la règle suivante :

SI on sort de la boucle parce que

$$\langle v^{k,n}, s^{k,n} \rangle > \sigma \|s^{k,n}\|^2 \quad (\text{ et } \ell \leq j_{k,n} + 1)$$

comme dans l'algorithme conceptuel,

ALORS $v^{k,n}$ définit un hyperplan suffisamment éloigné de x^k , ou, en d'autres mots, $\|x^{k+1} - x^k\|$ ne tend pas trop rapidement vers 0.

On obtient un nouveau point x^{k+1} comme dans l'algorithme conceptuel en projetant x^k sur H_{y^k, v^k} .

On parle alors de PAS SÉRIEUX.

SINON on sort de la boucle parce que $\ell = j_{k,n} + 1$ et

$$\langle v^{k,n}, s^{k,n} \rangle \leq \sigma \|s^{k,n}\|^2 ;$$

On obtient alors

$$\begin{aligned} \|y^{k,n} - x^k\| &= \left\| x^k - R 2^{-\ell} \frac{s^{k,n}}{\|s^{k,n}\|} - x^k \right\| \quad [\text{cf. définitions (2.d),(2.b),}] \\ &= |R 2^{-\ell}| \left\| \frac{s^{k,n}}{s^{k,n}} \right\| \\ &= R 2^{-j_{k,n}-1} \quad [\text{cf. } \ell = j_{k,n} + 1] \end{aligned}$$

et donc

$$\|y^{k,n} - x^k\| \leq R 2^{-j_{k,n}}$$

Dans ce cas,

- on n'incrmente pas k ;
- on ajoute la paire $(y^{k,n}, v^{k,n})$ au faisceau $\{(z^i, w^i)\}_{i \leq p}$, c'est-à-dire on pose $(z^{p+1}, w^{p+1}) = (y^{k,n}, v^{k,n})$;
- on incrmente n ;
- on calcule une nouvelle direction de recherche $s^{k,n+1}$ [tenant compte de $(y^{k,n}, v^{k,n})$ car le couple est tel que $\|y^{k,n} - x^k\| \leq R 2^{-j}$

et donc $(p+1) \in I_{k,n,j} = \{1 \leq i \leq p+1 \mid \|z^i - x^k\| \leq R 2^{-j}\}$ tel que ce $s^{k,n+1}$ est une meilleure approximation de l'élément de norme minimum de $T^{\varepsilon 2^{-j}}$.

On parle alors de PAS NUL.

On voit maintenant immédiatement que l'indice p nous donne le nombre d'éléments qui se trouvent dans le faisceau $\{(z^i, w^i \in T(z^i))\}_{i \leq p}$ et l'indice n nous donne à chaque itération le nombre d'approximations faites (c'est-à-dire de directions de recherche $s^{k,n}$ calculées), pour trouver la bonne paire $(y^{k,n}, v^{k,n})$, c'est-à-dire celle qui, dans le pas sérieux, nous fournit x^{k+1} en projetant x^k sur H_{y^k, v^k} .

△

Après cette présentation de la méthode faisceau et des généralisations nécessaires pour pouvoir l'adapter à notre cas, on peut maintenant écrire l'algorithme implémentable en bonne et due forme.

3.2 L'algorithme implémentable

Choisir les paramètres $\tau > 0$, $R > 0$, $\sigma \in]0, 1[$.

Poser $k := 0$ et $p := 0$. Prendre $x^0 \in \mathbb{R}^n$.

Pas # 0 : Test d'arrêt

(0.a) Calculer $u^k \in T(x^k)$.

Si $u^k = 0$, alors STOP.

(0.b) Sinon, $p := p + 1$,

$(z^p, w^p) := (x^k, u^k)$,

poser $n := 0$.

Pas # 1 : Calcul de la direction de recherche

(1.a) Poser $j := 0$;

(1.b) Définir $I_{k,n,j} = \{1 \leq i \leq p \mid \|z^i - x^k\| \leq R 2^{-j}\}$;

(1.c) Calculer $\alpha^{k,n,j} : = \operatorname{argmin} \left\{ \left\| \sum_{i \in I_{k,n,j}} \alpha_i w^i \right\|^2 \mid \sum_{i \in I_{k,n,j}} \alpha_i = 1 \right\}$;

(1.d) Prendre $s^{k,n,j} = \sum_{i \in I_{k,n,j}} \alpha_i^{k,n,j} w^i$;

(1.e) Si $\|s^{k,n,j}\| \leq \tau 2^{-j}$, alors : $j : = j + 1$,
aller en (1.b) .

(1.f) Sinon, définir $j_{k,n} : = j$,
 $s^{k,n} : = s^{k,n,j_{k,n}}$.

Pas # 2 : Recherche linéaire

(2.a) Poser $\ell : = 0$.

(2.b) Définir $y^{k,n,\ell} : = x^k - R 2^{-\ell} \frac{s^{k,n}}{\|s^{k,n}\|}$, et prendre $v^{k,n,\ell} \in T(y^{k,n,\ell})$.

(2.c) Si $\langle v^{k,n,\ell}, s^{k,n} \rangle \leq \sigma \|s^{k,n}\|^2$ ET $\ell < j_{k,n} + 1$, alors : $\ell : = \ell + 1$,
aller en (2.b) .

(2.d) Sinon, définir $\ell_{k,n} : = \ell$,
 $y^{k,n} : = y^{k,n,\ell_{k,n}}$,
 $v^{k,n} : = v^{k,n,\ell_{k,n}}$.

Pas # 3 : Évaluation de la paire (y, v)

(3.a) [PAS NUL]

Si $\langle v^{k,n}, s^{k,n} \rangle \leq \sigma \|s^{k,n}\|^2$, alors : $p : = p + 1$,

$(z^p, w^p) : = (y^{k,n}, v^{k,n})$,

$n : = n + 1$,

aller en (1.b) .

(3.b) [PAS SÉRIEUX]

Sinon : définir

$$\begin{aligned}
n_k &:= n, \\
j_k &:= j_{k,n_k}, \\
\ell_k &:= \ell_{k,n_k}, \\
s^k &:= s^{k,n_k}, \\
y^k &:= y^{k,n_k}, \\
v^k &:= v^{k,n_k};
\end{aligned}$$

$$\text{définir } x^{k+1} = x^k - \frac{\langle v^k, x^k - y^k \rangle}{\|v^k\|^2} v^k;$$

$$\text{poser } k := k + 1;$$

aller en (0.a) .

Remarques :

1. Au **Pas # 0**, au lieu de vérifier si $u^k = 0$, on peut aussi vérifier si $\|u^k\| \leq \varepsilon$ (où ε est une tolérance fixée). Les preuves de convergence restent essentiellement les mêmes.
2. Au **Pas # 1**, pour $z^i = x^k$, on a toujours $0 \leq R 2^{-j}$; le vecteur $u^k \in T(x^k)$ se trouve donc toujours dans le sous-faisceau défini par les indices $I_{k,n,j}$. Cet ensemble n'est donc pas vide et $\alpha^{k,n,j}$ est bien défini.
3. Comme annoncé, le faisceau $\{(z^i, w^i \in T(z^i))\}_{i \leq p}$ contient à chaque instant toute l'information dont on dispose à ce moment-là: les points sérieux sont ajoutés en (0.b); les points obtenus lors des pas nuls sont ajoutés en (3.a) .
4. Au **Pas # 3**, si on est en (3.b), on a toujours les résultats suivants:

$$(a) \ell_k \leq j_k + 1$$

En effet, par les définitions (3.b),

$$\begin{aligned}
\ell_k &:= \ell_{k,n_k}, \\
j_k &:= j_{k,n_k}, \\
n_k &:= n,
\end{aligned}$$

et

$$\ell_{k,n} \leq j_{k,n} + 1$$

car pour arriver en (3.b) , il faut sortir de la boucle en (2.c) avec $\ell \leq j_{k,n} + 1$ [cf. test (3.a)].

$$(b) \langle v^k, s^k \rangle > \sigma \|s^k\|^2$$

En effet, par les définitions (3.b) et (2.d) :

$$\begin{aligned} n_k &:= n , \\ v^k &:= v^{k,n_k} = v^{k,n} : = v^{k,n,\ell_{k,n}} , \\ s^k &:= s^{k,n_k} = s^{k,n} \end{aligned}$$

et

$$\langle v^{k,n,\ell_{k,n}}, s^{k,n} \rangle > \sigma \|s^{k,n}\|^2$$

car on est en (3.b) , donc le test en (3.a) est faux, ce qui signifie aussi qu'on est sorti de la boucle en (2.c) avec l'inégalité précédente.

$$(c) \|s^k\| > \tau 2^{-j_k}$$

En effet, par les définitions (3.b) et (1.f) :

$$\begin{aligned} n_k &:= n , \\ s^k &:= s^{k,n_k} = s^{k,n} : = s^{k,n,j_{k,n}} , \\ j_k &:= j_{k,n_k} = j_{k,n} \end{aligned}$$

et

$$\|s^{k,n,j_{k,n}}\| > \tau 2^{-j_{k,n}}$$

car, pour arriver en (3.b) , il faut être sorti de la boucle en (1.e) .

$$(d) y^k = x^k - R 2^{-\ell_k} \frac{s^k}{\|s^k\|}$$

En effet, par les définitions (3.b) et (2.d) :

$$\begin{aligned} n_k &:= n , \\ y^k &:= y^{k,n_k} = y^{k,n} : = y^{k,n,\ell_{k,n}} , \\ s^k &:= s^{k,n_k} = s^{k,n} , \\ \ell_k &:= \ell_{k,n_k} = \ell_{k,n} \end{aligned}$$

et

$$y^{k,n,\ell_{k,n}} = x^k - R 2^{-\ell_{k,n}} \frac{s^{k,n}}{\|s^{k,n}\|}$$

par définition en (2.b) .

Avant de faire une analyse de convergence détaillée, on a besoin d'un lemme technique préliminaire.

Lemme 3.1

Soit $\gamma > 0$ fixé.

Soient deux suites infinies $\{t^p\}$ et $\{\tilde{t}^p\}$ telles que

$$\langle t^i - t^{p+1}, \tilde{t}^p \rangle \geq \gamma \|\tilde{t}^p\|^2, \quad \forall i = 1, \dots, p,$$

et $\{t^p\}$ est bornée.

Alors, $\tilde{t}^p \rightarrow 0$ ($p \rightarrow +\infty$).

Preuve : En appliquant l'inégalité de Cauchy-Schwarz à

$$\langle t^i - t^{p+1}, \tilde{t}^p \rangle \geq \gamma \|\tilde{t}^p\|^2, \quad \forall i = 1, \dots, p,$$

on obtient

$$\begin{aligned} \gamma \|\tilde{t}^p\|^2 &\leq \|\tilde{t}^p\| \|t^i - t^{p+1}\|, & i = 1, \dots, p, \\ \|\tilde{t}^p\| &\leq \frac{1}{\gamma} \|t^i - t^{p+1}\|, & i = 1, \dots, p. \end{aligned}$$

Supposons par l'absurde qu'il existe une sous-suite telle que

$$0 < \delta < \|\tilde{t}^{p_q}\| \leq \frac{1}{\gamma} \|t^i - t^{p_q+1}\|, \quad i = 1, \dots, p_q.$$

Alors, en extrayant de la sous-suite (bornée car $\{t^p\}$ est bornée) correspondante une sous-suite convergente, on obtient pour un i précédent $p_q + 1$ dans cette suite

$$0 < \delta < \frac{1}{\gamma} \|t^i - t^{p_q+1}\| \rightarrow 0,$$

ce qui est impossible. ■

3.3 Analyse de convergence

C'est dans l'analyse des pas nuls que réside la principale différence dans l'analyse de convergence entre l'algorithme conceptuel et l'algorithme implémentable.

Dans ce dernier, on a quatre boucles respectivement sur les indices k , j , ℓ et n . On va montrer que, ou bien l'algorithme s'arrête après un nombre fini d'étapes, et alors le

dernier point est optimal, ou bien il génère une suite infinie de points qui converge vers un point de \mathcal{S} (boucle k). Si x^k n'est pas optimal, on n'aura jamais de boucle infinie pour les indices j , ℓ et n .

Plus précisément,

Proposition 3.1

Soit x^k l'itéré courant, non optimal, de l'algorithme implémentable.

Alors, on a les résultats suivants :

1. *il existe $j = j_{k,n}$ fini tel qu'on quitte la boucle en j (1.b)-(1.e) avec*

$$\|s^{k,n}\| > \tau 2^{-j_{k,n}} .$$

2. *il existe $\ell = \ell_{k,n}$ fini tel qu'on quitte la boucle en ℓ (2.b)-(2.c) avec*

$$\ell_{k,n} \leq j_{k,n} + 1 .$$

3. *il existe $n = n_{k,n}$ fini tel qu'on quitte la boucle en n (1.b)-(3.a) avec*

$$\langle v^{k,n}, s^{k,n} \rangle > \sigma \|s^{k,n}\|^2 .$$

Remarque : cette proposition est l'équivalent de la Proposition 2.2 pour l'algorithme conceptuel.

Preuve :

1. Supposons par l'absurde que l'algorithme boucle à l'infini en j ; alors on aurait une suite infinie $\{s^{k,n,j}\}_{j \in \mathbb{N}}$ telle que

$$\|s^{k,n,j}\| \leq \tau 2^{-j} .$$

On peut donc trouver deux sous-suites $\{n_q\}$ et $\{j_q\}$ telles que

$$\|s^{k,n_q,j_q}\| \leq \tau 2^{-j_q} \tag{3.1}$$

avec

$$\lim_{q \rightarrow +\infty} j_q = +\infty .$$

En effet, si on regarde la boucle pour un n fixé, il suffit de prendre la suite constante

$$n_q \equiv n , \quad \forall q \quad \text{et} \quad \{j_q\} = \{j\} .$$

Si, par contre, on tient compte du fait qu'en (3.a), on incrémente n , puis on revient en (1.b), on pourrait a priori avoir une suite

$$\{n_q\}_q \rightarrow +\infty \quad (q \rightarrow +\infty);$$

on prendra alors, par exemple,

$$\{j_q\}_q = \{j_{k,n_q} - 1\}_q.$$

Par (1.b), on a

$$\forall i \in I_{k,n_q,j_q} : \|z^i - x^k\| \leq R 2^{-j_q}.$$

Considérons ensuite les combinaisons convexes suivantes

$$\begin{aligned} \hat{x}^q &= \sum_{i \in I_{k,n_q,j_q}} \alpha_i^{k,n_q,j_q} z^i, \\ \hat{s}^q &= \sum_{i \in I_{k,n_q,j_q}} \alpha_i^{k,n_q,j_q} w^i = s^{k,n_q,j_q} \quad [\text{cf. (1.d)}] \end{aligned}$$

(où α_i^{k,n_q,j_q} est donné par (1.c)).

En appliquant le Corollaire 1.3 avec

$$\begin{aligned} \tilde{x} &= x^k, & \rho &= R 2^{-j_q}, & (\text{on a bien } \|z^i - x^k\| \leq R 2^{-j_q}) \\ \tilde{z} &= \hat{x}^q, & \hat{w} &= \hat{s}^q, \end{aligned}$$

on obtient

$$\|\hat{x}^q - x^k\| \leq R 2^{-j_q}, \quad (3.2)$$

et

$$\hat{s}^q \in T^{\hat{\varepsilon}_q}(\hat{x}^q)$$

avec

$$\hat{\varepsilon}_q \leq 2 R 2^{-j_q} M \quad (3.3)$$

où

$$M = \max\{\|w^i\| \mid i \in I_{k,n_q,j_q}\},$$

or

$$\begin{aligned} w^i &= \begin{cases} u^k \in T(x^k), \\ \text{ou} \\ v^{k,n,\ell} \in T(y^{k,n,\ell}) = T\left(x^k - R 2^{-\ell} \frac{s^{k,n}}{\|s^{k,n}\|}\right), \quad \ell \geq 0, \end{cases} \\ \Rightarrow w^i &\in T\left(\overline{B(x^k, R)}\right), \end{aligned}$$

d'où

$$M = \sup\{\|u\| \mid u \in T(\overline{B(x^k, R)})\} .$$

En passant à la limite pour $q \rightarrow +\infty$,

$$(3.1) \quad \text{donne} \quad \hat{s}^q = s^{k,n_q,j_q} \rightarrow 0 ,$$

$$(3.2) \quad \text{donne} \quad \hat{x}^q \rightarrow x^k ,$$

$$(3.3) \quad \text{donne} \quad \hat{\varepsilon}^q \rightarrow 0 ,$$

Comme $\hat{s}^q \in T^{\hat{\varepsilon}^q}(\hat{x}^q)$, la Proposition 1.5 donne

$$0 \in T^0(x^k) = T(x^k) ,$$

ce qui est impossible car par hypothèse, x^k n'est pas optimale.

La boucle en j est donc finie et on a démontré 1.

2. On ne boucle jamais à l'infini en ℓ , car le test en (2.c) sera nécessairement une fois faux : on incrémente ℓ chaque fois qu'on a

$$\langle v^{k,n,\ell}, s^{s,k} \rangle \leq \sigma \|s^{k,n}\|^2 \quad \text{et} \quad \ell < j_{k,n} + 1 ,$$

Alors, ou bien on finira par atteindre $j_{k,n} + 1$ et le test sera faux à l'étape suivante, ou bien on tombe sur un $v^{k,n,\ell}$ qui est tel que

$$\langle v^{k,n,\ell}, s^{k,n} \rangle > \sigma \|s^{k,n}\|^2 .$$

Dans les deux cas, on sort de la boucle avec

$$\ell_{k,n} \leq j_{k,n} + 1 .$$

3. Supposons par l'absurde que l'algorithme boucle à l'infini en n ; alors on aurait une suite infinie $\{(y^{k,n}, v^{k,n} \in T(y^{k,n}))\}_{n \in \mathbb{N}}$.

On vient de prouver en 1. que

- pour chaque n , la boucle en j se termine; donc $s^{k,n} = s^{k,n,j_{k,n}}$ est telle que

$$\|s^{k,n}\| > \tau 2^{-j_{k,n}} . \tag{3.4}$$

- $\{j_{k,n}\}_{n \in \mathbb{N}}$ est bornée, car sinon ce serait une sous-suite de j tendant vers l'infini, ce qui est impossible.

Comme, dans la boucle (3.a)-(1.b), j n'est jamais remis à 0, on a une suite croissante bornée, donc convergente, de nombres entiers.

Il existe donc un rang \bar{n} à partir duquel

$$j_{k,n} = j_{k,\bar{n}}, \quad \forall n \geq \bar{n}.$$

Considérons maintenant la suite $\{(y^{k,n}, v^{k,n})\}_{n \geq \bar{n}}$ à partir de ce rang.

Or, dire qu'on a une boucle infinie en n revient à dire qu'on fait une infinité de pas nuls, c'est-à-dire

$$\langle v^{k,n}, s^{k,n} \rangle \leq \sigma \|s^{k,n}\|^2 \quad (3.5)$$

et

$$\begin{aligned} \|y^{k,n} - x^k\| &= \left\| x^k - R 2^{-\ell_{k,n}} \frac{s^{k,n}}{\|s^{k,n}\|} - x^k \right\| \\ &= R 2^{-\ell_{k,n}} \\ &= R 2^{-j_{k,n}-1} \quad [\text{cf. pas nul : on est donc sorti de la boucle en} \\ &\quad (2.c) \text{ parce que } \ell_{k,n} = j_{k,n} + 1] \\ &= R 2^{-j_{k,\bar{n}}-1}, \quad \forall n \geq \bar{n}, \\ &\leq R 2^{-j_{k,\bar{n}}}, \quad \forall n \geq \bar{n}. \end{aligned}$$

Par conséquent, on satisfait (1.b) et la paire $(y^{k,n}, v^{k,n})$ est ajoutée au sous-faisceau [cf. explication pages 59–60] défini par les indices $I_{k,n,j_{k,n}}$ et ceci, pour tout $n \geq \bar{n}$.

Soit alors un indice \tilde{n} tel que $\bar{n} < \tilde{n} < n$.

Comme $j_{k,n} = j_{k,\bar{n}}$ pour tout $n \geq \bar{n}$, alors j n'augmente plus, c'est-à-dire qu'on ne renforce plus l'inégalité dans

$$I_{k,n,j} = \{1 \leq i \leq p \mid \|z^i - x^k\| \leq R 2^{-j}\}$$

et, par conséquent, $(y^{k,\tilde{n}}, v^{k,\tilde{n}})$ reste dans le sous-faisceau définissant l'élément de norme minimum pour tout $n > \tilde{n}$.

Donc, par la caractérisation de la projection P d'un point x sur un ensemble E convexe fermé,

$$\forall y \in E : \quad \langle x - P(x), y - P(x) \rangle \leq 0,$$

donc, pour $x = 0$, $y = v^{k,\tilde{n}}$:

$$\langle 0 - s^{k,n}, v^{k,\tilde{n}} - s^{k,n} \rangle \leq 0,$$

$$\langle v^{k,\tilde{n}}, s^{k,n} \rangle \geq \|s^{k,n}\|^2, \quad \forall n > \tilde{n} > \bar{n}. \quad (3.6)$$

(3.6)-(3.5) donne

$$\langle v^{k,\tilde{n}} - v^{k,n}, s^{k,n} \rangle \geq (1 - \sigma) \|s^{k,n}\|^2, \quad \forall n > \tilde{n} > \bar{n},$$

ou encore, en posant

$$\left. \begin{aligned} t^i &\equiv v^{k,\bar{n}+i}, \\ \tilde{t}^i &\equiv s^{k,(\bar{n}+1)+i}, \end{aligned} \right\} \quad i \geq 1,$$

on obtient

$$\langle t^i - t^{p+1}, \tilde{t}^p \rangle \geq (1 - \sigma) \|\tilde{t}^p\|^2, \quad \forall p > i > 1.$$

De plus, $\{t^i\}$ est bornée car

$$\begin{aligned} t^i &= v^{k,\bar{n}+i}, & i \geq 1, \\ &= v^{k,n}, & n \geq \bar{n}, \end{aligned}$$

et

$$v^{k,n} \in T(y^{k,n}) \quad \text{avec } y^{k,n} = x^k - R 2^{-\ell} \frac{s^{k,n}}{\|s^{k,n}\|},$$

c'est-à-dire

$$y^{k,n} \in B(x^k, R)$$

et donc $\{v^{k,n}\}_{n > \bar{n}} \subseteq T(B(x^k, R))$ est bornée car T étant maximal monotone est localement borné [cf. Proposition 1.7].

On satisfait donc les hypothèses du Lemme 3.1 avec $\gamma = (1 - \sigma)$.

On peut donc conclure que

$$\begin{aligned} \tilde{t}^p &\rightarrow 0 & (p \rightarrow +\infty), \\ \text{c'est-à-dire } s^{k,n} &\rightarrow 0 & (n \rightarrow +\infty). \end{aligned}$$

Or, on sait par (3.4) que

$$\begin{aligned} \|s^{k,n}\| &> \tau 2^{-j_{k,n}} \\ &> \tau 2^{-j_{k,\bar{n}}} > 0, & \forall n \geq \bar{n}, \end{aligned}$$

ce qui est impossible.

La boucle en n est donc finie et on a démontré 3. ■

La proposition suivante analyse toutes les possibilités qui peuvent se présenter lors d'une itération de l'algorithme implémentable.

Proposition 3.2

Soit x^k l'itéré courant de l'algorithme implémentable.

Alors, les situations suivantes se présentent :

1. *Si x^k est optimal,*

alors ou bien la sous-routine qui calcule $u^k \in T(x^k)$ donne $u^k = 0$ et l'algorithme s'arrête en (0.a) ;

ou bien l'algorithme boucle indéfiniment après ce dernier pas sérieux sans mettre à jour k .

2. *si x^k n'est pas optimal,*

alors l'algorithme atteint (3.b) après un nombre fini d'itérations internes.

De plus,

$$\|s^{k,n_k^*,j_k-1}\| \leq \tau 2^{-j_k+1}$$

où n_k^ est la plus petite valeur de n égalant $j_{k,n} = j_k$ pour tout $j_k > 0$.*

Preuve :

1. Si x^k est optimal et si la sous-routine donne $u^k = 0$, le test d'arrêt en (0.a) est satisfait et, par conséquent, l'algorithme s'arrête.

Si, par contre, $u^k \neq 0$, supposons par l'absurde qu'on atteint (3.b) et qu'on effectue un nouveau pas sérieux qui donne x^{k+1} .

On aurait alors

$$\langle x^k - y^k, v^k \rangle = \left\langle R 2^{-\ell_k} s^k \frac{1}{\|s^k\|}, v^k \right\rangle$$

[cf. Remarque 4 (d) qui suit la description de l'algorithme]

et ensuite successivement

$$\begin{aligned}
\langle x^k - y^k, v^k \rangle &= R 2^{-\ell_k} \frac{1}{\|s^k\|} \langle s^k, v^k \rangle \\
&> R 2^{-\ell_k} \frac{1}{\|s^k\|} \sigma \|s^k\|^2 \quad [\text{cf. Remarque 4 (b) ibidem}] \\
&> R 2^{-\ell_k} \tau 2^{-j_k} \sigma \quad [\text{cf. Remarque 4 (c) ibidem}] \\
&> 0 .
\end{aligned}$$

Or, comme T est monotone et que

$$\begin{aligned}
0 &\in T(x^k) \quad [\text{cf. } x^k \text{ est optimal}] \\
v^k &\in T(y^k) ,
\end{aligned}$$

on a, par la Définition 1.5,

$$\begin{aligned}
\langle 0 - v^k, x^k - y^k \rangle &\geq 0 , \\
-\langle v^k, x^k - y^k \rangle &\geq 0 , \\
\langle x^k - y^k, v^k \rangle &\leq 0 ,
\end{aligned}$$

ce qui contredit l'inégalité précédente.

2. Si x^k n'est pas optimal, la Proposition 3.1 montre qu'il n'y a pas de boucle infinie en j , ℓ et n à l'intérieur d'une itération k , et que donc (3.b) est finalement atteint.

Pour démontrer l'inégalité, posons

$$n_k^* = \min\{n \leq n_k \mid j_{k,n} = j_k\} ;$$

alors les indices

$$j = j_{k,n_k^*} - 1 < j_k ,$$

et

$$j + 1 = j_{k,n_k^*} = j_k ,$$

dans la boucle en (1.b)-(1.e) sont tels que (1.e) est vrai pour l'indice j et (1.f) est vrai pour l'indice $j + 1$, c'est-à-dire

$$\begin{aligned}
\|s^{k,n_k^*,j_{k,n_k^*}-1}\| &\leq \tau 2^{-(j_{k,n_k^*}-1)} \\
\|s^{k,n_k^*,j_{k,n_k^*}}\| &> \tau 2^{-j_{k,n_k^*}} ,
\end{aligned}$$

on a donc bien

$$\|s^{k,n_k^*,j_k-1}\| \leq \tau 2^{-j_k+1} . \quad \blacksquare$$

Cette Proposition 3.2 montre que, comme pour l'algorithme conceptuel [cf. Proposition 2.2], ou bien l'algorithme implémentable s'arrête après un nombre fini d'étapes et alors le dernier point est optimal, ou bien l'algorithme implémentable génère une suite infinie de points dont aucun n'est optimal.

Or, par le Théorème 2.1, on sait que la suite $\{x^k\}$ générée par l'algorithme conceptuel est ou bien finie et le dernier élément est optimal, ou bien elle est infinie et elle converge vers une solution du problème :

$$\text{trouver } x \text{ tel que } 0 \in T(x) .$$

On aimerait donc avoir le même résultat de convergence pour la suite infinie générée par l'algorithme implémentable. Avant de pouvoir prouver cette convergence [cf. Théorème 3.1], il nous faut quelques lemmes technique préliminaires.

Le premier résultat technique est l'équivalent de la Proposition 2.3 :

Proposition 3.3

Soit x^k l'itéré courant (non optimal) de l'algorithme implémentable.

Alors, après avoir défini x^{k+1} en (3.b) , on a

1. $x^k \notin H_{y^k,v^k} \stackrel{\text{def}}{=} \{z \in \mathbb{R}^n \mid \langle z - y^k, v^k \rangle \leq 0\}$
 $x^{k+1} = \text{Proj}_{H_{y^k,v^k}}(x^k)$
2. $\forall x^* \in \mathcal{S} : \|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 - \|x^{k+1} - x^k\|^2$
3. $\|x^{k+1} - x^k\| > R \sigma \tau 2^{-2j_k-1} \frac{1}{\|v^k\|}$

Preuve :

1. Dans la preuve de la Proposition 3.2, 1., on a montré que le couple (y^k, v^k) est tel que

$$\langle x^k - y^k, v^k \rangle > R 2^{-\ell_k-j_k} \tau \sigma > 0$$

d'où $x^k \notin H_{y^k,v^k}$

$$x^{k+1} = \text{Proj}_{H_{y^k,v^k}}(x^k)$$

car par la définition en (3.b) ,

$$x^{k+1} := x^k - \frac{\langle v^k, x^k - y^k \rangle v^k}{\|v^k\|^2} .$$

2. est immédiat comme dans la preuve de la proposition 2.3, 2.

3. Par définition de x^{k+1} ,

$$\begin{aligned} \|x^{k+1} - x^k\| &= \left\| -\frac{\langle v^k, x^k - y^k \rangle v^k}{\|v^k\|^2} \right\| \\ &= \frac{|\langle v^k, x^k - y^k \rangle| \|v^k\|}{\|v^k\|^2} . \end{aligned}$$

En reprenant la même minoration qu'au point 1., on obtient immédiatement

$$\begin{aligned} \|x^{k+1} - x^k\| &> R \sigma \tau 2^{-\ell_k - j_k} \frac{1}{\|v^k\|} \\ &> R \sigma \tau 2^{-2j_k - 1} \frac{1}{\|v^k\|} \end{aligned}$$

[cf. Remarque 4 (a) qui suit la description de l'algorithme :

$$\ell_k \leq j_k + 1 \Rightarrow -\ell_k \geq -j_k - 1 .] \blacksquare$$

Le deuxième résultat technique est l'équivalent du Lemme 2.1.

Lemme 3.2

Si l'algorithme implémentable génère une suite infinie $\{x^k\}$,

alors elle converge au sens de Fejer vers l'ensemble des solutions \mathcal{S} .

De plus, x^k , s^k , y^k et v^k , de même que les ensembles $\{s^{k,n,j}\}$, $\{(y^{k,n}, v^{k,n})\}$ et $\{(z^p, w^p)\}$ sont bornés.

Preuve : Remarquons que, par définition,

$$s^k = s^{k,n_k} = s^{k,n_k,j_k,n_k} ,$$

$$y^k = y^{k,n_k} ,$$

$$v^k = v^{k,n_k} .$$

Il suffit donc de montrer que les ensembles sont bornés car, alors, s^k , y^k et v^k le seront aussi puisque ce sont des éléments de ces ensembles.

Par la Proposition 3.3, 2., on a

$$\begin{aligned} \forall x^* \in \mathcal{S} : \quad \|x^{k+1} - x^*\|^2 &\leq \|x^k - x^*\|^2 - \underbrace{\|x^{k+1} - x^k\|^2}_{\geq 0} \\ &\leq \|x^k - x^*\|^2 \end{aligned}$$

d'où, par définition 2.1, $\{x^k\}$ converge au sens de Fejer vers l'ensemble \mathcal{S} .

En vertu de la Proposition 2.1, 1., la suite $\{x^k\}$ est alors également bornée, c'est-à-dire

$$\exists c \text{ tel que } x_k \leq c, \quad \forall c.$$

Donc, il existe un compact K_0 (prendre par exemple $K_0 = \overline{B(0, c)}$) tel que

$$\{x^k\} \subseteq K_0. \quad (3.7)$$

En ce qui concerne $\{(y^{k,n}, v^{k,n})\}$, on a par les définitions (2.d) et (2.b),

$$\begin{aligned} y^{k,n} &= y^{k,n,\ell_{k,n}} \\ &= x^k - R 2^{-\ell_{k,n}} \frac{s^{k,n}}{\|s^{k,n}\|} \end{aligned}$$

Comme $\{x^k\} \subseteq K_0$ [cf. (3.7)], $\ell_{k,n} \geq 0$ et $\left\| \frac{s^{k,n}}{\|s^{k,n}\|} \right\| = 1$,

on a, en définissant $K_1 \equiv K_0 + \overline{B(0, R)}$,

$$y^{k,n} \in K_1.$$

Comme T est maximal monotone, il est localement borné et donc borné sur des parties bornées en vertu des Propositions 1.7 et 1.6.

Par conséquent, $T(\{y^{k,n}\}) = \{v^{k,n}\}$ est borné de même que $T(\{x^k\}) = \{u^k\}$.

En ce qui concerne $\{(z^p, w^p)\}$, z^p est soit extrait de $\{x^k\}$ en (0.b), soit de $\{y^k\}$ en (3.a).

Dans les deux cas, $z^p \in K_1$ et est donc borné.

w^p est borné car extrait des ensembles bornés $\{u^k\}$ en (0.b), ou $\{v^{k,n}\}$ en (3.a).

Finalement, $\{s^{k,n,j}\}$ est borné car, par (1.d) ,

$$s^{k,n,j} = \sum_{i \in I_{k,n,j}} \alpha_i^{k,n,j} w^i$$

est une somme convexe de w^p qui sont bornés. ■

Le dernier résultat technique avant la preuve de convergence de l'algorithme jouera le rôle du Lemme 3.2.

Lemme 3.3

*Si l'algorithme implémentable génère une suite infinie $\{x^k\}$, alors $\lim_{k \rightarrow +\infty} j_k = +\infty$.
En d'autres mots, les indices j et k tendent ensemble vers l'infini.*

Preuve : Par la Proposition 3.3, 2., on a les résultats suivants :

$$\forall x^* \in \mathcal{S} : \quad \|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 - \|x^{k+1} - x^k\|^2$$

c'est-à-dire

$$\begin{aligned} \|x^1 - x^0\|^2 &\leq \|x^0 - x^*\|^2 - \|x^1 - x^*\|^2 \\ &\vdots \\ \|x^{n+1} - x^n\|^2 &\leq \|x^n - x^*\|^2 - \|x^{n+1} - x^*\|^2 . \end{aligned}$$

En sommant,

$$\begin{aligned} \sum_{k=0}^n \|x^{k+1} - x^k\|^2 &\leq \|x^0 - x^*\|^2 - \underbrace{\|x^{n+1} - x^*\|^2}_{\geq 0} \\ &\leq \|x^0 - x^*\|^2 , \end{aligned}$$

en faisant tendre $n \rightarrow +\infty$, on obtient

$$\sum_{k=0}^{+\infty} \|x^{k+1} - x^k\|^2 < +\infty$$

car le membre de droite est un nombre fini indépendant de n .

La condition nécessaire de convergence nous permet alors d'affirmer que

$$\lim_{k \rightarrow +\infty} \|x^{k+1} - x^k\| = 0 . \tag{3.8}$$

Par la Proposition 3.3, 3., on sait que

$$\begin{aligned} \|x^{k+1} - x^k\| &> R \sigma \tau 2^{-2j_k-1} \frac{1}{\|v^k\|} \\ &> \frac{R \sigma \tau}{M} 2^{-2j_k-1} \end{aligned}$$

[cf. v^k est borné en vertu du Lemme 3.2, on note M une borne supérieure.]

En laissant tendre $k \rightarrow +\infty$, on obtient ainsi successivement

$$0 \geq \lim_{k \rightarrow +\infty} \frac{R \sigma \tau}{M} 2^{-2j_k-1} \geq 0 \quad [\text{cf. (3.8)}]$$

$$0 \geq \lim_{k \rightarrow +\infty} 2^{-2j_k-1} \geq 0 \quad [\text{cf. } R, \tau, M > 0, \sigma \in]0, 1[]$$

$$\lim_{k \rightarrow +\infty} 2^{-2j_k-1} = 0$$

$$\lim_{k \rightarrow +\infty} j_k = +\infty . \quad \blacksquare$$

Remarque : cette preuve est analogue à la 1ère partie de la preuve du Théorème 2.1. La seule différence est que là, on obtient

$$\lim_{k \rightarrow +\infty} \ell_k + 2 j_k = +\infty$$

et il faut utiliser le Lemme 2.2 pour conclure que j_k n'est pas bornée, alors qu'ici, on obtient immédiatement

$$\lim_{k \rightarrow +\infty} j_k = +\infty .$$

Ceci provient du fait que, dans l'algorithme implémentable, on peut se servir du test supplémentaire en (2.c) :

$$\ell_k \leq j_k + 1$$

pour obtenir une minoration de $\|x^{k+1} - x^k\|$ dépendant uniquement de j_k , plus de ℓ_k [cf. Propositions 2.3 et 3.3].

Comme annoncé à la page 72, on peut maintenant prouver l'équivalent du Théorème 2.1.

Théorème 3.1

Considérons la suite $\{x^k\}$ générée par l'algorithme implémentable.

Alors

- ou bien, elle est finie, et le dernier élément est optimal;*
- ou bien elle est infinie, et elle converge vers une solution du problème :*

trouver x tel que $0 \in T(x)$.

Preuve :

- Le cas fini a été traité avec la Proposition 3.2.
- Si la suite est infinie, par le Lemme 3.2, elle converge au sens de Fejer; par la Proposition 2.1, 1., elle est alors bornée et on peut donc extraire une sous-suite convergente (au sens de la norme) :

$$\lim_{q \rightarrow +\infty} x^{k_q} \equiv \hat{x} .$$

Comme, par le Lemme 3.3, on a

$$\lim_{k \rightarrow +\infty} j_k = +\infty ,$$

on a également

$$\lim_{q \rightarrow +\infty} j_{k_q} = +\infty$$

et on peut supposer que, pour q suffisamment grand,

$$j_{k_q} > 0 .$$

Alors, en posant

$$n_k^* = \min\{n \leq n_k \mid j_{k,n} = j_k\} ,$$

la Proposition 3.2 donne

$$0 \leq \|s^{k_q, n_{k_q}^*, j_{k_q}-1}\| \leq \tau 2^{-j_{k_q}+1} . \quad (3.9)$$

Par (1.b) , on a

$$\forall i \in I_{k_q, n_{k_q}^*, j_{k_q}-1} : \|z^i - x^{k_q}\| \leq R 2^{-j_{k_q}+1} .$$

Comme dans la preuve de la Proposition 3.1, 1., considérons ensuite les combinaisons convexes suivantes :

$$\begin{aligned}\hat{x}^q &= \sum_{i \in I_{k_q, n_{k_q}^*, j_{k_q}-1}} \alpha_i^{k_q, n_{k_q}^*, j_{k_q}-1} z^i \\ \hat{s}^q &= \sum_{i \in I_{k_q, n_{k_q}^*, j_{k_q}-1}} \alpha_i^{k_q, n_{k_q}^*, j_{k_q}-1} w^i = s^{k_q, n_{k_q}^*, j_{k_q}-1} \quad [\text{cf. (1.d) }]\end{aligned}$$

(où $\alpha_i^{k_q, n_{k_q}^*, j_{k_q}-1}$ est donné par (1.c)).

En appliquant le Corollaire 1.3 avec

$$\begin{aligned}\tilde{x} &= x^{k_q}, \quad \rho = R 2^{-j_{k_q}+1}, \quad (\text{on a bien } \|z^i - x^{k_q}\| \leq R 2^{-j_{k_q}+1}) \\ \hat{z} &= \hat{x}^q, \quad \hat{w} = \hat{s}^q,\end{aligned}$$

on obtient

$$\|\hat{x}^q - x^{k_q}\| \leq R 2^{-j_{k_q}+1}, \quad (3.10)$$

et

$$\hat{s}^q \in T^{\hat{\varepsilon}_q}(\hat{x}^q)$$

avec

$$\hat{\varepsilon}_q \leq 2 R 2^{-j_{k_q}+1} M \quad (3.11)$$

où

$$M = \max\{\|w^i\| \mid i \in I_{k_q, n_{k_q}^*, j_{k_q}-1}\}$$

(ces variables sont bornées en vertu du Lemme 3.2).

En passant à la limite pour $q \rightarrow +\infty$,

$$(3.9) \quad \text{donne} \quad \hat{s}^q = s^{k_q, n_{k_q}^*, j_{k_q}-1} \rightarrow 0;$$

$$(3.10) \quad \text{donne} \quad \|\hat{x}^q - x^{k_q}\| \rightarrow 0 \text{ et donc } \hat{x}^q \rightarrow \hat{x} \text{ car } x^{k_q} \rightarrow \hat{x};$$

$$(3.11) \quad \text{donne} \quad \hat{\varepsilon}^q \rightarrow 0.$$

Comme $\hat{s}^q \in T^{\hat{\varepsilon}_q}(\hat{x}^q)$, la Proposition 1.5 donne

$$0 \in T^0(\hat{x}) = T(\hat{x}),$$

c'est-à-dire $\hat{x} \in \mathcal{S}$.

La suite $\{x^k\}$ convergente au sens de Fejer a donc un point d'adhérence appartenant à \mathcal{S} , et, par la Proposition 2.1, 2., toute la suite converge donc vers un point dans \mathcal{S} . ■

Chapitre 4

Méthode hybride extragradient-point proximal approchée

4.1 Introduction et motivation

Dans ce chapitre, on va développer une autre méthode pour résoudre le problème :

$$\text{trouver } x \in X \text{ tel que } 0 \in T(x)$$

où X est un espace de Hilbert réel, et

$T : X \rightarrow \mathcal{P}(X)$ un opérateur maximal monotone.

En transformant un tout petit peu le problème de départ :

$$\begin{aligned} 0 \in T(x^*) &\Leftrightarrow 0 \in c T(x^*) && [c > 0] \\ &\Leftrightarrow x^* \in x^* + c T(x^*) \\ &\Leftrightarrow x^* \in (I + c T)(x^*) \\ &\Leftrightarrow x^* \in (I + c T)^{-1}(x^*) \\ &\Leftrightarrow x^* \in J_c(x^*) \end{aligned}$$

(où J_c représente la résolvante de T , cf. p. 6.),

on tombe sur une classe de méthodes très connue pour résoudre ce problème : la **classe des méthodes du point proximal** (ou méthodes “prox”) dont l’idée est de trouver le point fixe de la résolvante de l’opérateur. Autrement dit, étant donné x^k , l’approximation courante

de la solution du problème, l'itéré suivant est obtenu en résolvant le sous-problème :

$$\begin{aligned} x^{k+1} \in (I + c T)^{-1}(x^k) &\Leftrightarrow x^k \in (I + c T)(x^{k+1}) \\ &\Leftrightarrow x^k \in x^{k+1} + c T(x^{k+1}) \\ &\Leftrightarrow 0 \in c T(x^{k+1}) + (x^{k+1} - x^k) . \end{aligned}$$

Rockafellar permet en plus à la constante c de varier d'une itération à l'autre. En d'autres mots, étant donnés un opérateur maximal monotone et une suite de scalaires positifs $\{c_k\}_{k \geq 0}$ appelés **paramètres de relaxation** (ou "grandeurs de pas"), on dit que la suite $\{x^k\}_{k \geq 0}$ est **générée par l'algorithme prox** si

$$\begin{aligned} 0 &\in c_k T(x^{k+1}) + (x^{k+1} - x^k) \\ \text{(ou encore } &x^{k+1} \in J_{c_k}(x^k) \text{)} . \end{aligned}$$

4.1.1 Le reste

Comme résoudre le problème précité peut être (presque) aussi difficile que résoudre le problème de départ, le théorème de convergence de Rockafellar permet d'évaluer approximativement les résolvantes, pourvu que la somme des erreurs soit finie.

Résoudre approximativement ces sous-problèmes peut s'écrire formellement

$$\begin{aligned} \text{trouver } &x^{k+1} \in X \quad \text{et} \quad v^{k+1} \in T(x^{k+1}) \\ \text{tels que } &0 = c_k v^{k+1} + (x^{k+1} - x^k) - r^k \end{aligned}$$

où $r^k \in X$ est le reste associé à la solution approchée du sous-problème qui doit satisfaire :

- $\|r^k\| \leq \sigma_k$ avec $\sum_{k=0}^{+\infty} \sigma_k < +\infty$ pour établir la convergence globale, et
- $\|r^k\| \leq \sigma_k \|x^{k+1} - x^k\|$ avec $\sum_{k=0}^{+\infty} \sigma_k < +\infty$ pour établir une vitesse de convergence linéaire locale.

De plus, $\{c_k\}$ doit être tel que

$$c_k \geq \bar{c} > 0, \quad \forall k .$$

[Pour les démonstrations de ces résultats, se référer à [13].]

Ces conditions que le reste doit vérifier pour pouvoir considérer x^k comme une solution approchée du sous-problème prox sont appelées **critères de tolérance d'erreur**.

Ceci implique que, si on choisit d'autres critères de tolérance d'erreur, on obtient une autre définition de solution approchée du sous-problème.

4.1.2 L' ε -élargissement

Dans l'article [2], Burachik, Iusem et Svaiter ont choisi une approche différente.

Au lieu d'introduire un terme reste, ils travaillent avec l' ε -élargissement T^ε de T [cf. Définition 1.10] pour définir une solution approchée du sous-problème prox.

En effet, ils considèrent la méthode prox généralisée exacte suivante :

étant donné x^k , l'itéré suivant est obtenu en résolvant le problème

$$0 \in c_k T(x^{k+1}) + \nabla f(x^{k+1}) - \nabla f(x^k)$$

où f est une fonction de régularisation adéquate.

(Notons que, pour $f(x) = \frac{1}{2} \|x\|^2$, on a $\nabla f(x) = x$ et on retombe sur la méthode prox classique. Il s'agit donc bien d'une généralisation.)

Ensuite, ils définissent une solution approchée de ce problème comme étant une solution **exacte** du problème :

$$0 \in c_k T^{\varepsilon_k}(x^{k+1}) + \nabla f(x^{k+1}) - \nabla f(x^k) \text{ pour une valeur adéquate de } \varepsilon_k .$$

Pour assurer la convergence, il faut alors imposer entre autres que

$$\sum_{k=0}^{+\infty} \varepsilon_k < +\infty .$$

On voit que cette condition est du même type que les critères de tolérance d'erreur décrits en 4.1.1.

Un inconvénient majeur de ce genre de conditions est que, très souvent, on n'a pas de règle constructive pour les obtenir : en effet, il existe une infinité de séries sommables et on ne spécifie pas comment choisir σ_k ou ε_k à l'itération k pour garantir la convergence. Pour pouvoir construire un algorithme implémentable, on voudrait donc avoir un critère

de tolérance d'erreur calculable, c'est-à-dire une condition sur le reste qui remplace la sommabilité a priori par quelque chose qu'on peut évaluer à chaque itération k de l'algorithme. C'est là une des principales motivations de ce chapitre.

4.1.3 Première approche

Dans l'article [14], Solodov et Svaiter ont proposé un nouveau critère de tolérance d'erreur qui a deux avantages :

1. il est précisément constructif;
2. le paramètre de relaxation σ (fixé) ne doit pas tendre vers zéro pour assurer la convergence.

Ce critère est décrit dans l'algorithme hybride projection-point proximal suivant :

- La paire (y^k, v^k) est admissible si et seulement si

$$v^k \in T(y^k) \\ \text{et} \quad 0 = c_k v^k + (y^k - x^k) - r^k$$

où r^k est le reste qui doit satisfaire

$$\|r^k\| \leq \sigma \max\{c_k \|v^k\|, \|y^k - x^k\|\}$$

avec $\sigma \in [0, 1[$ (fixé).

- L'itéré suivant est obtenu en projetant x^k sur l'hyperplan

$$\{z \in X \mid \langle v^k, z - y^k \rangle = 0\}$$

(dont on peut montrer qu'il sépare strictement x^k de chaque zéro de T).

$$S \subseteq \{z \mid \langle v^k, z - y^k \rangle \leq 0\}$$

$$\sigma \in [0, 1[$$

Cette méthode hybride préserve toutes les propriétés de convergence de la méthode prox exacte (respectivement de sa version approchée décrite en 4.1.1).

Remarquons encore qu'on peut donner une interprétation de ce critère en termes d'erreur relative. En effet, une solution exacte du sous-problème prox :

$$\text{trouver } x \text{ tel que } 0 \in c_k T(x) + x - x^k$$

$$\textcircled{X} \text{ If } \|r^k\| \leq \sigma c_k \|v^k\| \quad (v^k \neq 0) \quad \text{then } \langle v^k, x^k - y^k \rangle > 0$$

$$\langle v^k, x^k - y^k \rangle = c_k \|v^k\|^2 - \langle r^k, v^k \rangle > 0$$

$$\text{because } \langle r^k, v^k \rangle \leq \|r^k\| \|v^k\| \leq \sigma c_k \|v^k\|^2 \leq c_k \|v^k\|^2 \quad \text{if } 0 > 0$$

$$\text{If } \|r^k\| \leq \sigma \|y^k - x^k\| \quad \dots$$

peut être regardée comme une paire (y^k, v^k) telle que

$$v^k \in T(y^k) \quad \text{et} \quad 0 = c_k v^k + y^k - x^k$$

alors que, pour une solution approchée, on a

$$r^k = c_k v^k + (y^k - x^k) .$$

Ainsi, pour estimer l'erreur relative, il faut analyser les rapports

$$\frac{\|r^k\|}{\|c_k v^k\|} \quad \text{et} \quad \frac{\|r^k\|}{\|y^k - x^k\|} ,$$

et le critère

$$\|r^k\| \leq \sigma \max\{c_k \|v^k\| , \|y^k - x^k\|\}$$

exprime simplement le fait que la borne pour l'erreur relative peut être fixée à σ et ne doit pas tendre vers 0 .

4.1.4 Deuxième approche

Dans ce chapitre, Solodov et Svaiter ont combiné leurs idées de [2] et [14] et ont proposé une méthode plus facile dans le sens qu'on n'effectue plus de projection.

On propose en fait un nouveau critère de tolérance d'erreur et, par conséquent, une nouvelle définition de solution approchée, qui fait intervenir à la fois l' ε -élargissement T^{ε_k} [cf. 4.1.2] et un reste r^k [cf. 4.1.1] constructif [cf. 4.1.3] dans l'équation du sous-problème prox :

- une solution approchée est une paire (y^k, v^k) telle que

$$\begin{aligned} v^k &\in T^{\varepsilon_k}(y^k) \\ \text{et} \quad 0 &= c_k v^k + (y^k - x^k) - r^k \end{aligned}$$

où ε_k, r^k sont relativement petits par rapport à $\|y^k - x^k\|$, plus précisément

$$\|r^k\|^2 + 2 c_k \varepsilon_k \leq \sigma^2 \|y^k - x^k\|^2 ;$$

- L'itéré suivant est alors donné par

$$x^{k+1} = x^k - c_k v^k . \quad \left(\begin{array}{l} = x^k + y^k - x^k - r^k \\ = y^k - r^k \end{array} \right)$$

À chaque itération de la méthode, la direction v^k est donc obtenue par une évaluation approchée de l'opérateur T en un point intermédiaire y^k , c'est-à-dire on effectue un pas de type extragradient approché [cf. [9]] et ceci, en utilisant des informations obtenues par une solution approchée d'un sous-problème prox. D'où le nom : **méthode hybride extragradient-point proximal approchée**.

4.2 La méthode hybride extragradient-point proximal approchée

Dans cette section, on introduit formellement le critère de tolérance d'erreur et la méthode présentée en 4.1.4. On donne quelques propriétés des solutions approchées et, à l'aide d'un exemple, on attire l'attention sur certaines caractéristiques des solutions approchées et de la méthode.

Définition 4.1

Soient $x \in X$, $c > 0$, $\sigma \in [0, 1]$.

On dit que la paire (y, v) est une solution approchée du problème

$$0 \in c T(\cdot) + (\cdot - x)$$

avec une tolérance d'erreur σ si et seulement si, pour un $\varepsilon \geq 0$, on a

$$v \in T^\varepsilon(y),$$

$$c v + (y - x) - r = 0,$$

$$\|r\|^2 + 2 c \varepsilon \leq \sigma^2 \|y - x\|^2.$$

En vertu du Théorème 1.1, on sait que, à chaque itération k , le sous-problème prox

$$0 \in c_k T(x^{k+1}) + (x^{k+1} - x^k)$$

(ou en termes de résolvante $x^{k+1} \in J_{c_k}(x^k)$)

admet toujours une solution unique (car T est maximal monotone).

Cette solution exacte (avec $\varepsilon = 0$ et $r = 0$) est évidemment toujours une solution approchée pour tout $\sigma \in [0, 1]$ car

$$\begin{aligned} \|r\|^2 + 2 c \varepsilon &\leq \sigma^2 \|y - x\|^2 \\ 0 &\leq \sigma^2 \|y - x\|^2 \end{aligned}$$

est toujours vérifiée.

Le sous-problème prox admet donc toujours au moins une solution approchée !

Dans le cas où $\sigma = 0$ (pas de tolérance d'erreur), la dernière inégalité devient

$$\|r\|^2 + 2 c \varepsilon \leq 0$$

et $\|r\|^2 \geq 0$, $\varepsilon \geq 0$, $c > 0$ implique alors que $r = 0$ et $\varepsilon = 0$ et la seule paire à satisfaire la définition $v \in T(x)$ et $c v + y - x = 0$ est la solution exacte du sous-problème prox.

Cette définition de solution approchée et le sens de tolérance d'erreur sont donc assez naturels.

4.2.1 Comparaison avec 4.1.2

On va montrer, à l'aide d'un exemple, que la notion de solution approchée introduite par la définition précédente est plus large que celle décrite en 4.1.2 qui, pour

$$f(x) = \frac{1}{2} \|x\|^2$$

est équivalente à

$$\text{trouver } x^{k+1} \in X, \quad \varepsilon_k \geq 0 \quad \text{et} \quad v^{k+1} \in T^{\varepsilon_k}(x^{k+1})$$

$$\text{tels que } 0 = c_k v^{k+1} + (x^{k+1} - x^k)$$

car, en général, la relation

$$v \in T^\varepsilon(y) \quad \text{et} \quad c v + (y - x) - r = 0 \quad (r \neq 0)$$

n'implique pas qu'il existe $\varepsilon' \geq 0$ tel que

$$0 \in c T^{\varepsilon'}(y) + y - x.$$

Pour s'en convaincre, considérons l'exemple suivant :

Soient $X = \mathbb{R}^2$ et l'opérateur T une rotation de $\frac{\pi}{2}$ rad.

T est donc univoque et peut être représenté par la matrice

$$M = \begin{pmatrix} \cos \frac{\pi}{2} & -\sin \frac{\pi}{2} \\ \sin \frac{\pi}{2} & \cos \frac{\pi}{2} \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

$$\begin{aligned}
T : \mathbb{R}^2 &\rightarrow \mathbb{R}^2, \\
x &\mapsto T(x) = M x, \\
\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &\mapsto \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix}.
\end{aligned}$$

- T est monotone [cf. Définition 1.5] car, pour tout $x, y \in \mathbb{R}^2$:

$$\begin{aligned}
\langle x - y, T(x) - T(y) \rangle &= (x_1 - y_1, x_2 - y_2) \begin{pmatrix} -x_2 + y_2 \\ x_1 - y_1 \end{pmatrix} \\
&= (x_1 - y_1) (-x_2 + y_2) + (x_2 - y_2) (x_1 - y_1) \\
&= 0 \geq 0.
\end{aligned}$$

- T est maximal monotone [cf. Définition 1.6] car :

soit $(x, y) \in \mathbb{R}^2 \times \mathbb{R}^2$ tel que

$$\langle x - \xi, y - \eta \rangle \geq 0, \quad \forall \xi \in \mathbb{R}^2, \eta = M \xi,$$

c'est-à-dire tel que

$$(x_1 - \xi_1) (y_1 + \xi_2) + (x_2 - \xi_2) (y_2 - \xi_1) \geq 0, \quad \forall (\xi_1, \xi_2),$$

ou encore

$$x_1 y_1 + x_2 y_2 \geq \xi_1 (y_1 + x_2) - \xi_2 (x_1 - y_2), \quad \forall (\xi_1, \xi_2). \quad (4.1)$$

Il faut alors montrer que

$$y = M x$$

$$\text{c'est-à-dire } (y_1, y_2) = (-x_2, x_1).$$

Pour $\xi_1 = 0$, (4.1) donne

$$x_1 y_1 + x_2 y_2 \geq -\xi_2 (x_1 - y_2), \quad \forall \xi_2 \in \mathbb{R}.$$

Supposons par l'absurde que $y_2 < x_1$.

Alors, en faisant tendre $\xi_2 \rightarrow -\infty$, on obtient

$$\langle x, y \rangle \geq +\infty ,$$

ce qui est impossible.

De même, supposons par l'absurde que $y_2 > x_1$.

Alors, en faisant tendre $\xi_2 \rightarrow +\infty$, on obtient

$$\langle x, y \rangle \geq +\infty ,$$

ce qui est impossible.

D'où : $y_2 = x_1$.

De même, pour $\xi_2 = 0$, (4.1) donne

$$x_1 y_1 + x_2 y_2 \geq \xi_1 (y_1 + x_2), \quad \forall \xi_1 \in \mathbb{R}.$$

Supposons par l'absurde que $y_1 < -x_2$.

Alors, en faisant tendre $\xi_1 \rightarrow -\infty$, on obtient

$$\langle x, y \rangle \geq +\infty ,$$

ce qui est impossible.

De même, supposons par l'absurde que $y_1 > -x_2$.

Alors, en faisant tendre $\xi_1 \rightarrow +\infty$, on obtient

$$\langle x, y \rangle \geq +\infty ,$$

ce qui est impossible.

D'où : $y_1 = -x_2$.

Ainsi, $y = M x$ et T est maximal monotone.

- $T^\varepsilon \equiv T$ c'est-à-dire

$$\forall x \in \mathbb{R}^2 : \quad T^\varepsilon(x) = T(x) \quad \text{et ceci, pour tout } \varepsilon \geq 0$$

car, par la Définition 1.10,

$$\begin{aligned} u \in T^\varepsilon(x) &\Leftrightarrow \langle u - v, x - y \rangle \geq -\varepsilon, \quad \forall y \in \mathbb{R}^2, v = T(y) \\ &\Leftrightarrow \langle u - M y, x - y \rangle \geq -\varepsilon, \quad \forall y \in \mathbb{R}^2 \\ &\Leftrightarrow (u_1 + y_2)(x_1 - y_1) + (u_2 - y_1)(x_2 - y_2) \geq -\varepsilon, \quad \forall (y_1, y_2) \\ &\Leftrightarrow y_1(-u_1 - x_2) + y_2(x_1 - u_2) + u_1 x_1 + u_2 x_2 \geq -\varepsilon, \quad \forall (y_1, y_2). \end{aligned} \quad (4.2)$$

Pour $y_1 = 0$, (4.2) donne

$$y_2 (x_1 - u_2) \geq -\varepsilon - u_1 x_1 - u_2 x_2, \quad \forall y_2 \in \mathbb{R}.$$

Supposons par l'absurde que $u_2 < x_1$.

Alors, en faisant tendre $y_2 \rightarrow -\infty$, on obtient

$$-\infty \geq -\varepsilon - \langle u, x \rangle,$$

ce qui est impossible.

De même, supposons par l'absurde que $u_2 > x_1$.

Alors, en faisant tendre $y_2 \rightarrow +\infty$, on obtient

$$-\infty \geq -\varepsilon - \langle u, x \rangle,$$

ce qui est impossible.

D'où : $u_2 = x_1$.

De même, pour $y_2 = 0$, (4.2) donne

$$y_1 (-u_1 - x_2) \geq -\varepsilon - u_1 x_1 - u_2 x_2, \quad \forall y_1 \in \mathbb{R}.$$

Supposons par l'absurde que $u_1 < -x_2$.

Alors, en faisant tendre $y_1 \rightarrow -\infty$, on obtient

$$-\infty \geq -\varepsilon - \langle u, x \rangle,$$

ce qui est impossible.

De même, supposons par l'absurde que $u_1 > -x_2$.

Alors, en faisant tendre $y_1 \rightarrow +\infty$, on obtient

$$-\infty \geq -\varepsilon - \langle u, x \rangle,$$

ce qui est impossible.

D'où : $u_1 = -x_2$.

Ainsi,

$$\begin{aligned} u \in T^\varepsilon(x) &\Leftrightarrow (u_1, u_2) = (-x_2, x_1) \\ &\Leftrightarrow u = M x = T(x) \end{aligned}$$

et $T^\varepsilon = T$.

Grâce à cet exemple, on voit bien qu'en général, la relation

$$v \in T^\varepsilon(y) \quad \text{et} \quad c v + (y - x) = r \quad (r \neq 0)$$

n'implique pas qu'il existe $\varepsilon' \geq 0$ tel que

$$0 \in c T^{\varepsilon'}(y) + y - x.$$

En effet, ici, l'opérateur T est univoque, maximal monotone et

$$T^\varepsilon(y) = T(y), \quad \forall \varepsilon \geq 0.$$

Ainsi, il est immédiat que la relation

$$v = T(y) \quad \text{et} \quad c v + (y - x) = r \quad (r \neq 0)$$

n'implique pas que

$$0 = c T(y) + y - x.$$

Dans le cas particulier de cet exemple, on peut donc uniquement traiter des solutions exactes avec la méthode décrite en 4.1.2.

4.2.2 Comparaison avec 4.1.3

La notion de solution approchée de la section 4.1.3 fait uniquement intervenir un terme reste, mais pas d' ε -élargissement. Pour la comparer avec la notion de solution approchée de la définition 4.1, considérons donc le cas où $\varepsilon = 0$.

Le critère de tolérance d'erreur devient alors

$$\begin{aligned} v &\in T(y), \\ c v + (y - x) &= r, \end{aligned}$$

et

$$\|r\| \leq \sigma \|y - x\|,$$

ce qui est seulement un peu moins exigeant que le critère de tolérance d'erreur décrit en 4.1.3.

En effet, il est évident que, si le reste satisfait le critère

$$\|r\| \leq \sigma \max\{c \|v\|, \|y - x\|\},$$

$$\|y - x\| \leq \sigma \|v\|.$$

alors il satisfait également le critère

$$\|r\| \leq \sigma \|y - x\| .$$

Si, par contre, le reste satisfait

$$\|r\| \leq \sigma \|y - x\| ,$$

il faut imposer en plus que $\sigma < \frac{1}{2}$ pour retrouver l'autre critère.

En effet, on a alors

$$\begin{aligned} c \|v\| &= \|c v\| \\ &= \| - c v \| \\ &= \|y - x - r\| && [\text{cf. } c v + y - x = r] \\ &\geq \left| \|y - x\| - \|r\| \right| \\ &\geq \left| \|y - x\| - \sigma \|y - x\| \right| && [\text{par hypothèse, } \|r\| \leq \sigma \|y - x\|] \\ &= \left| (1 - \sigma) \|y - x\| \right| \\ &= (1 - \sigma) \|y - x\| ; && [\text{car } \|y - x\| \geq 0 \text{ et } \sigma \in [0, \frac{1}{2}[] \end{aligned}$$

d'où

$$\|r\| \leq \sigma \|y - x\| \leq \sigma \frac{c}{1 - \sigma} \|v\| .$$

En posant $\tilde{\sigma} \equiv \frac{\sigma}{1 - \sigma}$, on a bien $\sigma \leq \tilde{\sigma} < 1$ car on a imposé $\sigma \in [0, \frac{1}{2}[$!

On obtient ainsi

$$\|r\| \leq \tilde{\sigma} c \|v\| \quad \text{et} \quad \|r\| \leq \sigma \|y - x\| \leq \tilde{\sigma} \|y - x\| .$$

Donc,

$$\|r\| \leq \tilde{\sigma} \max\{c \|v\| , \|y - x\|\} .$$

(Notons que, même si $\tilde{\sigma} \geq \sigma$, on a quand même $\frac{\tilde{\sigma}}{\sigma} \approx 1$ pour $0 < \sigma \ll 1$.)

4.2.3 Propriété des solutions approchées

Dans la suite, si on parle de solutions approchées d'un sous-problème prox, on entend toujours par là des solutions satisfaisant la Définition 4.1.

La première question qui se pose est évidemment de savoir de combien une solution approchée d'un sous-problème prox diffère de sa solution exacte. La réponse est donnée par le Lemme suivant :

Lemme 4.1

Soient $x \in X$ et $c > 0$.

Soit (y, v) une solution approchée du problème

$$0 \in c T(\cdot) + (\cdot - x)$$

avec une tolérance d'erreur $\sigma \in [0, 1[$.

Soit z la solution exacte de ce problème.

Alors, on a

$$\|z - y\| \leq \sigma \|y - x\| .$$

Preuve : Par hypothèse, (y, v) est une solution approchée du problème

$$0 \in c T(\cdot) + (\cdot - x)$$

avec une tolérance d'erreur σ . D'où, par la Définition 4.1, il existe $\varepsilon \geq 0$ tel que

$$v \in T^\varepsilon(y) , \tag{4.3}$$

$$c v + (y - x) = r , \tag{4.4}$$

$$\|r\|^2 + 2 c \varepsilon \leq \sigma^2 \|y - x\|^2 . \tag{4.5}$$

Par hypothèse, z est la solution exacte du problème prox (qui existe toujours en vertu du Théorème 1.1 car T est maximal monotone), c'est-à-dire

$$0 \in c T(z) + (z - x)$$

$$\Leftrightarrow -\frac{z - x}{c} \in T(z) .$$

Comme, par (4.3), $v \in T^\varepsilon(y)$, on a, par la Définition 1.10,

$$\langle u - v, x - y \rangle \geq -\varepsilon , \quad \forall x \in X, \forall u \in T(x)$$

donc en particulier pour $x = z$ et $u = -\frac{z-x}{c} \in T(z)$, on a

$$\left\langle -\frac{z-x}{c} - v, z-y \right\rangle \geq -\varepsilon. \quad (4.6)$$

Or,

$$\begin{aligned} -\frac{z-x}{c} - v &= -\frac{z-x}{c} - \frac{1}{c} (r-y+x) && [\text{cf. (4.4)}] \\ &= -\frac{1}{c} (z-x+r-y+x) \\ &= -\frac{1}{c} r - \frac{1}{c} (z-y). \end{aligned}$$

Ainsi, en remplaçant dans (4.6), on obtient :

$$\begin{aligned} &\left\langle -\frac{1}{c} r - \frac{1}{c} (z-y), z-y \right\rangle \geq -\varepsilon \\ \Leftrightarrow &-\frac{1}{c} \langle r, z-y \rangle - \frac{1}{c} \|z-y\|^2 \geq -\varepsilon \\ \Leftrightarrow &\langle r, z-y \rangle + \|z-y\|^2 \leq c \varepsilon && [(-c) < 0] \\ \Leftrightarrow &-\|r\| \|z-y\| + \|z-y\|^2 \leq c \varepsilon && [\text{Cauchy-Schwarz}] \\ \Leftrightarrow &\|z-y\|^2 - \|r\| \|z-y\| - c \varepsilon \leq 0. \end{aligned}$$

Ce trinôme du second degré en $\|z-y\|$ a le signe du coefficient de $\|z-y\|^2$ (c'est-à-dire positif) sauf entre les racines qui sont données par

$$\frac{-(-\|r\|) \pm \sqrt{(-\|r\|)^2 - 4 \cdot 1 \cdot (-c \varepsilon)}}{2 \cdot 1}.$$

L'inégalité précédente implique donc que

$$\frac{\|r\| - \sqrt{\|r\|^2 + 4 c \varepsilon}}{2} \leq 0 \leq \|z-y\| \leq \frac{\|r\| + \sqrt{\|r\|^2 + 4 c \varepsilon}}{2}$$

d'où :

$$\|z-y\| \leq \frac{\|r\|}{2} + \frac{1}{2} \sqrt{\|r\|^2 + 4 c \varepsilon}$$

ou encore

$$\|z-y\| \leq \sqrt{\left(\frac{\|r\|}{2}\right)^2 + \frac{\|r\|^2}{4} + c \varepsilon}.$$

Comme la fonction $f : x \mapsto \sqrt{x}$ est concave, on a

$$f(x_1) + f(x_2) \leq f(x_1 + x_2) .$$

Ainsi

$$\begin{aligned} \|z - y\| &\leq \sqrt{\frac{\|r\|^2}{4} + \frac{\|r\|^2}{4} + c \varepsilon} \\ &= \sqrt{\frac{\|r\|^2}{2} + c \varepsilon} \\ &\leq \sqrt{2} \sqrt{\frac{\|r\|^2}{2} + c \varepsilon} && [\text{car } \sqrt{2} > 1] \\ &= \sqrt{\|r\|^2 + 2 c \varepsilon} \\ &\leq \sqrt{\sigma^2 \|y - x\|^2} && [\text{cf. la fonction } f \text{ est croissante et (4.5)}] \\ &= \sigma \|y - x\| ; && [\text{cf. } \sigma \in [0, 1[\text{ donc termes positifs ou nuls}] \end{aligned}$$

on a donc bien

$$\|z - y\| \leq \sigma \|y - x\| . \quad \blacksquare$$

Présentons maintenant formellement la méthode décrite en 4.1.4.

4.2.4 L'algorithme hybride extragradient-point proximal approché

Initialisation :

Choisir $x^0 \in X$,

$\sigma \in [0, 1[$,

$\bar{c} > 0$.

Itération : Pour $k = 0, 1, 2, \dots$

Pas # 1

Choisir $c_k \geq \bar{c}$ et trouver (y^k, v^k) comme une solution approchée de

$$0 \in c_k T(\cdot) + (\cdot - x^k)$$

avec une tolérance d'erreur σ , c'est-à-dire pour $\varepsilon_k \geq 0$,

$$v^k \in T^{\varepsilon_k}(y^k)$$

et

$$\|c_k v^k + (y^k - x^k)\|^2 + 2 c_k \varepsilon_k \leq \sigma^2 \|y^k - x^k\|^2.$$

Pas # 2

Définir $x^{k+1} := x^k - c_k v^k$.

4.2.5 Remarques

1. Le cas où $\sigma = 0$

Notons que, si le k -ième sous-problème prox est résolu exactement,

alors la k -ième itération de l'algorithme ci-dessus coïncide avec une itération de la méthode prox classique.

En effet, comme $\varepsilon_k = 0$, on a alors

$$v^k \in T(y^k)$$

et comme $c_k v^k + (y^k - x^k) = 0$, (c'est-à-dire le reste est nul), on a

$$\begin{aligned} x^{k+1} &= x^k - c_k v^k \\ &= y^k. \end{aligned}$$

Ainsi, x^{k+1} est choisi tel que

$$0 \in c_k T(x^{k+1}) + (x^{k+1} - x^k),$$

ce qui est bien une itération de la méthode prox classique [cf. page 81].

Comme dans le cas particulier où $\sigma = 0$, la solution exacte du sous-problème prox est la seule à satisfaire le Pas # 1 de l'algorithme [cf. commentaire page 86 qui suit la Définition 4.1], on retrouve dans ce cas la méthode prox exacte classique.

Dans la suite, on va donc surtout s'intéresser au cas où $\sigma > 0$; on travaille donc avec des itérations prox approchées.

2. Autre écriture de l'algorithme

Remarquons qu'on a décrit l'algorithme en termes de $\{x^k\}$, $\{(y^k, v^k)\}$ et $\{\varepsilon_k\}$ parce que c'est plus pratique pour l'implémenter et pour en donner des applications.

En effet, à chaque itération interne d'une méthode utilisée pour résoudre le k -ième sous-problème prox

$$0 \in c_k T(\cdot) + (\cdot - x^k) ,$$

les valeurs de y et v qui constituent l'approximation courante de la solution du sous-problème sont typiquement disponibles. Il est donc naturel d'exprimer le critère d'arrêt du sous-problème (c'est-à-dire la tolérance d'erreur) en termes de y et v .

Pourtant, on pourrait sans problème réécrire l'algorithme en termes de $\{x^k\}$, $\{r^k\}$ et $\{\varepsilon_k\}$ car

$$x^{k+1} = x^k - c_k v^k \quad \text{et} \quad c_k v^k + (y^k - x^k) = r^k$$

donnent immédiatement

$$v^k \in T^{\varepsilon_k}(y^k) \quad \Leftrightarrow \quad \frac{x^k - x^{k+1}}{c_k} \in T^{\varepsilon_k}(x^{k+1} + r^k)$$

et

$$\begin{aligned} & \|c_k v^k + (y^k - x^k)\|^2 + 2 c_k \varepsilon_k \leq \sigma^2 \|y^k - x^k\|^2 \\ \Leftrightarrow & \|r^k\|^2 + 2 c_k \varepsilon_k \leq \sigma^2 \|r^k + x^{k+1} - x^k\|^2 . \end{aligned}$$

3. Importance du pas "extragradient"

On va maintenant montrer, à l'aide d'un exemple, que le pas "extragradient" est essentiel pour la convergence de la méthode.

Autrement dit, la méthode prox approchée avec le même critère de tolérance d'erreur, c'est-à-dire notre algorithme avec le Pas # 2 remplacé par $x^{k+1} = y^k$, peut ne pas converger.

Pour cela, reprenons l'exemple du paragraphe 4.2.1.

Comme on sait que

$$T^\varepsilon \equiv T , \quad \forall \varepsilon \geq 0 ,$$

on peut prendre

$$\varepsilon_k = 0 , \quad \forall k ,$$

sans changer quoi que ce soit au résultat.

Une suite générée par l'algorithme correspondant est donc telle que

$$\begin{cases} w^{k+1} \in T(x^{k+1}), \\ \|c_k w^{k+1} + x^{k+1} - x^k\| \leq \sigma \|x^{k+1} - x^k\|. \end{cases} \quad [\text{cf. } \varepsilon_k = 0 \text{ et } x^{k+1} = y^k]$$

On va montrer qu'une suite vérifiant ces conditions peut diverger :

Soient $x^0 \neq (0, 0)$,

$$\sigma \in \left] \frac{\sqrt{2}}{2}, 1 \right[,$$

$$c_k = \frac{1}{2}, \forall k ,$$

$$Q = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} .$$

Définir $Q^k = \underbrace{Q \cdot Q \cdot \dots \cdot Q}_{k \text{ fois}}$,

$$x^k = Q^k x^0 ,$$

$$w^k = T(x^k) = M x^k .$$

La suite $\{x^k\}$ convient.

En effet, par définition,

$$\begin{aligned} x^{k+1} - x^k &= Q x^k - x^k \\ &= (Q - I) x^k ; \end{aligned}$$

donc

$$\begin{aligned} c_k w^{k+1} + x^{k+1} - x^k &= \frac{1}{2} M x^{k+1} + (Q - I) x^k \\ &= \left(\frac{1}{2} M Q + Q - I \right) x^k ; \end{aligned}$$

or

$$\begin{aligned} \frac{1}{2} M Q + Q - I &= \frac{1}{2} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} . \end{aligned}$$

Ainsi,

$$\begin{aligned} \|c_k w^{k+1} + x^{k+1} - x^k\| &= \left\| \left(\frac{1}{2} M Q + Q - I \right) x^k \right\| \\ &\leq \left\| \frac{1}{2} M Q + Q - I \right\| \|x^k\| \\ &\quad [\text{compatibilité d'une norme matricielle (avec } \|A\| = \sqrt{\rho(A^* A)} \text{)}] \\ &= \frac{\sqrt{2}}{2} \|x^k\|. \end{aligned}$$

[car la racine de la plus grande valeur propre de

$$\begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \text{ vaut } \frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2}]$$

De même,

$$\begin{aligned} \|x^k\| &= \|(Q - I)^{-1} (x^{k+1} - x^k)\| \\ &\leq \|(Q - I)^{-1}\| \|x^{k+1} - x^k\| \quad [\text{compatibilité}] \\ &= \|x^{k+1} - x^k\|; \end{aligned}$$

[car la racine de la plus grande valeur propre de

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ vaut } 1]$$

d'où

$$\begin{aligned} \|c_k w^{k+1} + x^{k+1} - x^k\| &\leq \frac{\sqrt{2}}{2} \|x^k\| \\ &\leq \frac{\sqrt{2}}{2} \|x^{k+1} - x^k\| \\ &\leq \sigma \|x^{k+1} - x^k\|. \quad [\text{cf. } \sigma \in] \frac{\sqrt{2}}{2}, 1[] \end{aligned}$$

La suite $\{x^k\}$ vérifie donc bien les conditions données plus haut mais on va montrer qu'elle n'est pas bornée et donc, qu'a fortiori, elle ne converge pas vers l'unique solution de

$$T(x) = 0 \quad \Leftrightarrow \quad \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \Leftrightarrow \quad x = 0 \quad \text{qui est l'origine.}$$

Ceci montre que le pas “extragradient” est indispensable pour la convergence si on utilise la Définition 4.1 de solution approchée pour résoudre les sous-problèmes prox.

Il reste à montrer que la suite $\{x^k\}$ n'est pas bornée : on a

$$\|Q^k\| = \sqrt{2^k}, \quad \forall k ;$$

par récurrence, pour $k = 0$:

$$Q^* Q = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 2^0 & 0 \\ 0 & 2^0 \end{pmatrix}$$

dont la racine de la plus grande valeur propre vaut $\sqrt{2^0}$.

Supposons que

$$(Q^k)^* Q^k = \begin{pmatrix} 2^k & 0 \\ 0 & 2^k \end{pmatrix} \quad \text{et donc} \quad \|Q^k\| = \sqrt{2^k}$$

et montrons que

$$(Q^{k+1})^* Q^{k+1} = \begin{pmatrix} 2^{k+1} & 0 \\ 0 & 2^{k+1} \end{pmatrix} \quad \text{et donc} \quad \|Q^{k+1}\| = \sqrt{2^{k+1}}.$$

Nous avons en effet

$$\begin{aligned} (Q^{k+1})^* Q^{k+1} &= (Q^k Q)^* Q^k Q \\ &= Q^* (Q^k)^* Q^k Q \\ &= Q^* \begin{pmatrix} 2^k & 0 \\ 0 & 2^k \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \quad [\text{par hypothèse de récurrence}] \\ &= \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 2^k & 2^k \\ -2^k & 2^k \end{pmatrix} \\ &= \begin{pmatrix} 2^{k+1} & 0 \\ 0 & 2^{k+1} \end{pmatrix} \end{aligned}$$

dont la racine de la plus grande valeur propre vaut $\sqrt{2^{k+1}}$.

Par suite,

$$\begin{aligned}
 \|x^k\| &= \|Q^k x^0\| && \text{[cf. définition de } x^k \text{]} \\
 &\leq \|Q^k\| \|x^0\| && \text{[compatibilité]} \\
 &= \sqrt{2^k} \|x^0\|. && \text{[par ce qui précède]}
 \end{aligned}$$

Comme, par hypothèse, $x^0 \neq (0, 0)$, on a $\|x^0\| \neq 0$
et on obtient immédiatement, en passant à la limite pour $k \rightarrow +\infty$,

$$\|x^k\| \rightarrow +\infty ;$$

par conséquent, $\{x^k\}$ n'est pas bornée.

4.3 Analyse de convergence

Le but de cette section est de montrer que la suite $\{x^k\}$ générée par l'algorithme converge faiblement vers une solution du problème :

$$\text{trouver } x \text{ tel que } 0 \in T(x)$$

s'il en admet une et n'est pas bornée sinon. On montre de plus que la vitesse de convergence est linéaire si T^{-1} est Lipschitz continu autour de zéro.

Avant de pouvoir montrer ces deux théorèmes fondamentaux, il nous faut comme toujours quelques résultats préliminaires.

Dans un premier lemme, on démontre une propriété principale des solutions approchées des sous-problèmes prox, qui est à la base de l'algorithme qu'on a décrit.

Lemme 4.2

*Soit x^k l'itéré courant de l'algorithme hybride extragradient-point proximal approché.
Soient $c_k > 0$ et (y^k, v^k) une solution approchée du problème*

$$0 \in c_k T(\cdot) + (\cdot - x^k)$$

avec une tolérance d'erreur $\sigma \in [0, 1[$.

Alors, après avoir défini

$$x^{k+1} = x^k - c_k v^k$$

au Pas # 2, on a

$$\forall x^* \in \mathcal{S} : \|x^* - x^k\|^2 - \|x^* - x^{k+1}\|^2 \geq (1 - \sigma^2) \|y^k - x^k\|^2 .$$

Preuve : On a

$$\begin{aligned} & \|x^* - x^k\|^2 - \|x^* - x^{k+1}\|^2 \\ &= \|x^* - y^k + y^k - x^k\|^2 - \|x^* - y^k + y^k - x^{k+1}\|^2 \\ &= \|x^* - y^k\|^2 + 2 \langle x^* - y^k, y^k - x^k \rangle + \|y^k - x^k\|^2 \\ &\quad - \|x^* - y^k\|^2 - 2 \langle x^* - y^k, y^k - x^{k+1} \rangle - \|y^k - x^{k+1}\|^2 \\ &= 2 \langle x^* - y^k, y^k - x^k - y^k + x^{k+1} \rangle + \|y^k - x^k\|^2 - \|y^k - x^{k+1}\|^2 \\ &= 2 \langle x^* - y^k, x^{k+1} - x^k \rangle + \|y^k - x^k\|^2 - \|y^k - x^{k+1}\|^2 \end{aligned} \quad (4.7)$$

Par hypothèse, (y^k, v^k) est une solution approchée du problème

$$0 \in c_k T(\cdot) + (\cdot - x^k)$$

avec une tolérance d'erreur σ ; d'où, par la Définition 4.1,

$$\exists \varepsilon_k \geq 0 \quad \text{tel que} \quad v^k \in T^{\varepsilon_k}(y^k) , \quad (4.8)$$

$$c_k v^k + (y^k - x^k) = r^k , \quad (4.9)$$

$$\|r^k\|^2 + 2 c_k \varepsilon_k \leq \sigma^2 \|y^k - x^k\|^2 . \quad (4.10)$$

Comme, par (4.8), $v^k \in T^{\varepsilon_k}(y^k)$, on a, par la Définition 1.10,

$$\langle u - v^k, x - y^k \rangle \geq -\varepsilon_k , \quad \forall x \in X , \forall u \in T(x) ;$$

donc, en particulier pour $x = x^*$ et $u = 0 \in T(x^*)$ [cf. $x^* \in \mathcal{S}$], on a

$$\langle 0 - v^k, x^* - y^k \rangle \geq -\varepsilon_k$$

et, en multipliant par $c_k > 0$,

$$\begin{aligned} c_k \langle x^* - y^k, 0 - v^k \rangle &\geq -c_k \varepsilon_k \\ \Leftrightarrow \langle x^* - y^k, -c_k v^k \rangle &\geq -c_k \varepsilon_k \\ \Leftrightarrow \langle x^* - y^k, x^{k+1} - x^k \rangle &\geq -c_k \varepsilon_k . \quad [\text{cf. } x^{k+1} = x^k - c_k v^k] \end{aligned}$$

En combinant cette inégalité avec (4.7), on obtient

$$\begin{aligned}
\|x^* - x^k\|^2 - \|x^* - x^{k+1}\|^2 &\geq -2 c_k \varepsilon_k + \|y^k - x^k\|^2 - \|y^k - x^{k+1}\|^2 \\
&= -2 c_k \varepsilon_k + \|y^k - x^k\|^2 - \|y^k - x^k + c_k v^k\|^2 \\
&\quad \text{[cf. définition de } x^{k+1} \text{]} \\
&= -2 c_k \varepsilon_k + \|y^k - x^k\|^2 - \|r^k\|^2 \quad \text{[cf. (4.9)]} \\
&= \|y^k - x^k\|^2 - (\|r^k\|^2 + 2 c_k \varepsilon_k) \\
&\geq (1 - \sigma^2) \|y^k - x^k\|^2 . \quad \text{[cf. (4.10)]} \blacksquare
\end{aligned}$$

Notons en passant que, si $\sigma = 0$, c'est-à-dire dans le cas de la méthode prox exacte classique [cf. 4.2.5, Remarque 1], $y^k = x^{k+1}$ et le lemme s'écrit

$$\begin{aligned}
\forall x^* \in \mathcal{S} : \quad &\|x^* - x^k\|^2 - \|x^* - x^{k+1}\|^2 \geq \|x^{k+1} - x^k\|^2 \\
&\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 - \|x^{k+1} - x^k\|^2 ,
\end{aligned}$$

ce qui est exactement l'inégalité des Propositions 2.3, 2., et 3.3, 2., utilisées déjà dans les preuves de convergence des algorithmes des Chapitres 2 et 3.

Corollaire 4.1

Si le problème

$$\text{trouver } x \text{ tel que } 0 \in T(x)$$

admet une solution,

alors les suites $\{x^k\}$, $\{(y^k, v^k)\}$ et $\{\varepsilon_k\}$ générées par l'algorithme hybride extragradient-point proximal approché sont telles que

1. $\{x^k\}$ est bornée.
2. (a) $\sum_{k=0}^{+\infty} \|y^k - x^k\|^2 < +\infty$.
(b) $\|y^k - x^k\| \xrightarrow{k \rightarrow +\infty} 0$.
3. $\{y^k\}$ est bornée.

$$\begin{aligned}
4. \quad (a) \quad & \sum_{k=0}^{+\infty} \|v^k\|^2 \leq (1 + \sigma)^2 \sum_{k=0}^{+\infty} \frac{\|y^k - x^k\|^2}{c_k^2} < +\infty . \\
(b) \quad & \|v^k\| \xrightarrow{k \rightarrow +\infty} 0 . \\
5. \quad (a) \quad & \sum_{k=0}^{+\infty} \varepsilon_k < +\infty . \\
(b) \quad & \varepsilon_k \xrightarrow{k \rightarrow +\infty} 0 .
\end{aligned}$$

Preuve : Par hypothèse, le problème

trouver x tel que $0 \in T(x)$

admet une solution.

Soit donc $x^* \in \mathcal{S} \neq \emptyset$.

1. En vertu du Lemme 4.2,

$$\|x^* - x^k\|^2 - \|x^* - x^{k+1}\|^2 \geq (1 - \sigma^2) \|y^k - x^k\|^2 .$$

Comme $\sigma \in [0, 1[$, on a

$$(1 - \sigma^2) \geq 0$$

et vu que $\|\cdot\| \geq 0$, on en déduit que

$$\|x^* - x^k\| \geq \|x^* - x^{k+1}\| .$$

Il s'ensuit que la suite $\{\|x^k - x^*\|\}$ est décroissante et bornée inférieurement par zéro; elle converge donc et, par conséquent, la suite $\{x^k\}$ est bornée.

2. (a) De nouveau, par le Lemme 4.2 appliqué à $k = 0, 1, \dots, q$:

$$\begin{aligned}
\|x^* - x^0\|^2 - \|x^* - x^1\|^2 &\geq (1 - \sigma^2) \|y^0 - x^0\|^2 , \\
\|x^* - x^1\|^2 - \|x^* - x^2\|^2 &\geq (1 - \sigma^2) \|y^1 - x^1\|^2 , \\
&\vdots \\
\|x^* - x^q\|^2 - \|x^* - x^{q+1}\|^2 &\geq (1 - \sigma^2) \|y^q - x^q\|^2 .
\end{aligned}$$

En sommant, on obtient

$$\|x^* - x^0\|^2 - \|x^* - x^{q+1}\|^2 \geq (1 - \sigma^2) \sum_{k=0}^q \|y^k - x^k\|^2 ;$$

d'où

$$\sum_{k=0}^q \|y^k - x^k\|^2 \leq \|x^* - x^0\|^2 . \quad [\text{cf. } \|x^* - x^{q+1}\|^2 \geq 0]$$

$$\text{et } \sigma \in [0, 1[\Rightarrow 0 < 1 - \sigma^2 \leq 1]$$

Comme le membre de droite est indépendant de q , il suffit de passer à la limite pour $q \rightarrow +\infty$ pour obtenir

$$\sum_{k=0}^{+\infty} \|y^k - x^k\|^2 < +\infty .$$

(b) Par 2.(a), on a

$$\sum_{k=0}^{+\infty} \|y^k - x^k\|^2 < +\infty .$$

Ainsi, la condition nécessaire de convergence implique que

$$\|y^k - x^k\| \xrightarrow{k \rightarrow +\infty} 0 .$$

3. Par 2.(b), on a

$$\|y^k - x^k\| \xrightarrow{k \rightarrow +\infty} 0$$

et, par 1., $\{x^k\}$ est bornée.

D'où $\{y^k\}$ est bornée.

4. (a) On sait, par le Pas # 1 de l'algorithme que

$$\forall k : \quad \|c_k v^k + (y^k - x^k)\|^2 + 2 c_k \varepsilon_k \leq \sigma^2 \|y^k - x^k\|^2$$

d'où

$$\forall k : \quad \|c_k v^k + (y^k - x^k)\|^2 \leq \sigma^2 \|y^k - x^k\|^2 \quad [\text{cf. } c_k > 0 \text{ et } \varepsilon_k \geq 0]$$

et

$$\forall k : \quad \|c_k v^k + (y^k - x^k)\| \leq \sigma \|y^k - x^k\| . \quad (4.11)$$

De plus,

$$\begin{aligned}
 c_k \|v^k\| &= \|c_k v^k\| \\
 &= \|c_k v^k + (y^k - x^k) - (y^k - x^k)\| \\
 &\leq \|c_k v^k + (y^k - x^k)\| + \|y^k - x^k\| \quad [\text{inégalité triangulaire}] \\
 &\leq \sigma \|y^k - x^k\| + \|y^k - x^k\|, \quad [\text{cf. (4.11)}]
 \end{aligned}$$

c'est-à-dire

$$\|v^k\| \leq \frac{1+\sigma}{c_k} \|y^k - x^k\|, \quad \forall k.$$

En élevant au carré et en sommant sur $k = 0, 1, \dots, q$, on obtient

$$\sum_{k=0}^q \|v^k\|^2 \leq (1+\sigma)^2 \sum_{k=0}^q \frac{\|y^k - x^k\|^2}{c_k^2}.$$

En passant à la limite pour $q \rightarrow \infty$ et en tenant compte de 2.(a) et du fait que

$$c_k \geq \bar{c} > 0, \quad \forall k,$$

il s'ensuit

$$\sum_{k=0}^{+\infty} \|v^k\|^2 \leq (1+\sigma)^2 \sum_{k=0}^{+\infty} \frac{\|y^k - x^k\|^2}{c_k^2} < +\infty.$$

(b) Par 4.(a), on a

$$\sum_{k=0}^{+\infty} \|v^k\|^2 < +\infty.$$

Ainsi, la condition nécessaire de convergence implique que $\|v^k\| \xrightarrow{k \rightarrow +\infty} 0$.

5. (a) On sait, par le Pas #1 de l'algorithme que

$$\forall k : \quad \|c_k v^k + (y^k - x^k)\|^2 + 2 c_k \varepsilon_k \leq \sigma^2 \|y^k - x^k\|^2;$$

d'où

$$\forall k : \quad 2 c_k \varepsilon_k \leq \sigma^2 \|y^k - x^k\|^2 \quad [\text{car } \|c_k v^k + y^k - x^k\|^2 \geq 0]$$

$$\forall k : \quad \varepsilon_k \leq \frac{1}{2} \sigma^2 \frac{\|y^k - x^k\|^2}{c_k}.$$

En sommant sur $k = 0, 1, \dots, q$, on obtient

$$\sum_{k=0}^q \varepsilon_k \leq \frac{1}{2} \sigma^2 \sum_{k=0}^q \frac{\|y^k - x^k\|^2}{c_k}.$$

En passant à la limite pour $q \rightarrow +\infty$ et en tenant compte de 2.(a) et du fait que

$$c_k \geq \bar{c} > 0, \quad \forall k,$$

il s'ensuit

$$\sum_{k=0}^{+\infty} \varepsilon_k \leq \frac{1}{2} \sigma^2 \sum_{k=0}^{+\infty} \frac{\|y^k - x^k\|^2}{c_k} < +\infty.$$

(b) Par 5.(a), on a

$$\sum_{k=0}^{+\infty} \varepsilon_k < +\infty.$$

Ainsi, la condition nécessaire de convergence implique que $\varepsilon_k \xrightarrow{k \rightarrow +\infty} 0$. ■

Corollaire 4.2

Si le problème

trouver x tel que $0 \in T(x)$

admet une solution,

alors tout point d'accumulation faible de la suite $\{x^k\}$ générée par l'algorithme hybride extragradient-point proximal approché, est une solution du problème.

Preuve : En vertu du Corollaire 4.1, 1., la suite $\{x^k\}$ est bornée. Elle admet donc au moins une sous-suite convergente pour la topologie faible $\sigma(X, X')$ (c'est-à-dire la topologie la moins fine rendant continue chaque $f \in X'$). Autrement dit, elle admet au moins un point d'accumulation faible.

Soient \bar{x} un tel point d'accumulation faible de $\{x^k\}$ et $\{x^{k_j}\}$ une sous-suite de $\{x^k\}$ convergent faiblement vers \bar{x} :

$$x^{k_j} \rightharpoonup \bar{x}.$$

Par le Corollaire 4.1, 2.(b), on sait que

$$\|y^k - x^k\| \xrightarrow{k \rightarrow +\infty} 0.$$

Par conséquent,

$$\|y^{k_j} - x^{k_j}\| \xrightarrow{j \rightarrow +\infty} 0,$$

(car si une suite converge, toute sous-suite converge vers la même limite)
et donc

$$y^{k_j} \rightharpoonup \bar{x} . \quad (4.12)$$

Par le Corollaire 4.1, 4.(b), on sait que

$$\|v^k\| \xrightarrow{k \rightarrow +\infty} 0 .$$

Par conséquent, $v^k \rightarrow 0$ et donc

$$v^{k_j} \rightarrow 0 . \quad (4.13)$$

Par le Corollaire 4.1, 5.(b), on sait que

$$\varepsilon_k \xrightarrow{k \rightarrow +\infty} 0 ,$$

et donc

$$\varepsilon_{k_j} \rightarrow 0 . \quad (4.14)$$

Comme, par le Pas # 1 de l'algorithme, $v^k \in T^{\varepsilon_k}(y^k)$ et donc

$$v^{k_j} \in T^{\varepsilon_{k_j}}(y^{k_j}) ,$$

on a, par la Définition 1.10,

$$\begin{aligned} \langle u - v^{k_j}, x - y^{k_j} \rangle &\geq -\varepsilon_{k_j} , \quad \forall x \in X , \forall u \in T(x) , \\ \langle u - 0, x - y^{k_j} \rangle &\geq \langle v^{k_j}, x - y^{k_j} \rangle - \varepsilon_{k_j} , \quad \forall x \in X , \forall u \in T(x) . \end{aligned} \quad (4.15)$$

En passant à la limite pour $j \rightarrow +\infty$ et en tenant compte de la propriété suivante dans les espaces de Hilbert :

$$\begin{aligned} t_n \rightharpoonup t &\Leftrightarrow \forall f \in X' : f(t_n) \rightarrow f(t) , \\ &\Leftrightarrow \forall z \in X : \langle t_n, z \rangle \rightarrow \langle t, z \rangle , \end{aligned}$$

[théorème de représentation de Riesz]

on trouve

$$\begin{aligned} \bullet \quad y^{k_j} \rightharpoonup \bar{x} \text{ [cf. (4.12)]} &\Leftrightarrow \forall z \in X \quad \langle y^{k_j}, z \rangle \rightarrow \langle \bar{x}, z \rangle \\ \text{d'où } \langle u - 0, x - y^{k_j} \rangle &\rightarrow \langle u - 0, x - \bar{x} \rangle . \end{aligned}$$

- $\langle v^{k_j}, x - y^{k_j} \rangle = \langle v^{k_j}, x - y^{k_j} + \bar{x} - \bar{x} \rangle$
 $= \langle v^{k_j}, x \rangle - \langle v^{k_j}, y^{k_j} \rangle + \langle v^{k_j}, \bar{x} \rangle - \langle v^{k_j}, \bar{x} \rangle$
 où $\langle v^{k_j}, x \rangle \rightarrow 0$ et $\langle v^{k_j}, \bar{x} \rangle \rightarrow 0$ car, par (4.13), $v^{k_j} \rightarrow 0$;
 de plus, $\langle v^{k_j}, y^{k_j} \rangle - \langle v^{k_j}, \bar{x} \rangle \rightarrow 0$,
 car, par (4.12), $y^{k_j} \rightharpoonup \bar{x} \Leftrightarrow \forall z \in X : \langle y^{k_j}, z \rangle \rightarrow \langle \bar{x}, z \rangle$
 d'où, pour $z = v^{k_j}$, $\langle v^{k_j}, y^{k_j} \rangle \rightarrow \langle v^{k_j}, \bar{x} \rangle$.
 Par conséquent, $\langle v^{k_j}, x - y^{k_j} \rangle \rightarrow 0$.

- $\varepsilon_{k_j} \rightarrow 0$ par (4.14) .

Ainsi, (4.15) donne

$$\langle u - 0, x - \bar{x} \rangle \geq 0 - 0 , \quad \forall x \in X , \forall u \in T(x) .$$

Comme T est maximal monotone, il s'ensuit par la Définition 1.6 que $0 \in T(\bar{x})$, c'est-à-dire $\bar{x} \in \mathcal{S}$. ■

On peut maintenant prouver les deux résultats principaux de convergence annoncés à la page 100.

Théorème 4.1

Si le problème

trouver x tel que $0 \in T(x)$

admet une solution, c'est-à-dire si $\mathcal{S} \neq \emptyset$,

alors la suite $\{x^k\}$ générée par l'algorithme hybride extragradient-point proximal approché converge faiblement vers une solution du problème.

Si le problème n'admet pas de solution, c'est-à-dire si $\mathcal{S} = \emptyset$,

alors la suite $\{x^k\}$ n'est pas bornée.

Preuve :

1er cas : $\mathcal{S} \neq \emptyset$

En vertu du Corollaire 4.1, 1., la suite $\{x^k\}$ est bornée; elle admet donc au moins une sous-suite convergente pour la topologie faible $\sigma(X, X')$ (c'est-à-dire la topologie la moins fine rendant continue chaque $f \in X'$). Autrement dit, elle admet au moins un point d'accumulation faible.

Soient \bar{x} un tel point d'accumulation faible de $\{x^k\}$ et $\{x^{k_j}\}$ une sous-suite de $\{x^k\}$ convergeant faiblement vers \bar{x} :

$$x^{k_j} \rightharpoonup \bar{x}.$$

Par le Corollaire 4.2, \bar{x} est une solution du problème

$$\text{trouver } x \text{ tel que } 0 \in T(x),$$

c'est-à-dire si $\bar{x} \in \mathcal{S}$.

Il suffit maintenant de montrer que \bar{x} est le seul point d'accumulation faible de $\{x^k\}$ car alors, toute sous-suite convergeant faiblement converge vers la même limite et la convergence faible de toute suite $\{x^k\}$ vers \bar{x} suit par le caractère borné.

Soit donc \hat{x} un deuxième point d'accumulation faible de $\{x^k\}$ et montrons que $\hat{x} \equiv \bar{x}$.

Comme \hat{x} est un point d'accumulation faible, on sait qu'il existe une sous-suite $\{x^{k_t}\}$ telle que

$$x^{k_t} \rightharpoonup \hat{x}$$

et, par le Corollaire 4.2, on sait que $\hat{x} \in \mathcal{S}$.

En vertu du Lemme 4.2,

$$\forall x^* \in \mathcal{S} : \quad \|x^* - x^k\|^2 - \|x^* - x^{k+1}\|^2 \geq (1 - \sigma^2) \|y^k - x^k\|^2.$$

Comme $\sigma \in [0, 1[$, on a $(1 - \sigma^2) \geq 0$ et, vu que $\|\cdot\| \geq 0$, on en déduit que

$$\forall x^* \in \mathcal{S} : \quad \|x^* - x^k\| \geq \|x^* - x^{k+1}\|.$$

Il s'ensuit que les suites $\{\|\bar{x} - x^k\|\}$ et $\{\|\hat{x} - x^k\|\}$ sont décroissantes [cf. $\bar{x}, \hat{x} \in \mathcal{S}$] et bornées inférieurement par zéro, donc elles convergent:

$$\lim_{k \rightarrow +\infty} \|\bar{x} - x^k\| \stackrel{\text{not}}{=} \bar{\beta},$$

$$\lim_{k \rightarrow +\infty} \|\hat{x} - x^k\| \stackrel{\text{not}}{=} \hat{\beta}.$$

De plus,

$$\begin{aligned}\|\bar{x} - x^{k_\ell}\|^2 &= \|\bar{x} - \hat{x} + \hat{x} - x^{k_\ell}\|^2 \\ &= \|\bar{x} - \hat{x}\|^2 + \|\hat{x} - x^{k_\ell}\|^2 + 2 \langle \bar{x} - \hat{x}, \hat{x} - x^{k_\ell} \rangle .\end{aligned}\quad (4.16)$$

En passant à la limite pour $\ell \rightarrow +\infty$ et en tenant compte de la propriété suivante dans les espaces de Hilbert :

$$\begin{aligned}t_n \rightharpoonup t &\Leftrightarrow \forall f \in X' : f(t_n) \rightarrow f(t) , \\ &\Leftrightarrow \forall z \in X : \langle t_n, z \rangle \rightarrow \langle t, z \rangle ,\end{aligned}$$

[théorème de représentation de Riesz]

on trouve

- $\|\bar{x} - x^{k_\ell}\| \rightarrow \bar{\beta}$
car, si une suite converge, toute sous-suite converge vers la même limite;
- $\|\hat{x} - x^{k_\ell}\| \rightarrow \hat{\beta}$
car, si une suite converge, toute sous-suite converge vers la même limite;
- $x^{k_\ell} \rightharpoonup \hat{x} \Leftrightarrow \forall z \in X : \langle x^{k_\ell}, z \rangle \rightarrow \langle \hat{x}, z \rangle$
d'où $\langle \bar{x} - \hat{x}, \hat{x} - x^{k_\ell} \rangle \rightarrow \langle \bar{x} - \hat{x}, \hat{x} - \hat{x} \rangle = 0$;

ainsi (4.16) donne

$$\bar{\beta}^2 = \|\bar{x} - \hat{x}\|^2 + \hat{\beta}^2 . \quad (4.17)$$

De même,

$$\begin{aligned}\|\hat{x} - x^{k_j}\|^2 &= \|\hat{x} - \bar{x} + \bar{x} - x^{k_j}\|^2 \\ &= \|\hat{x} - \bar{x}\|^2 + \|\bar{x} - x^{k_j}\|^2 + 2 \langle \hat{x} - \bar{x}, \bar{x} - x^{k_j} \rangle .\end{aligned}\quad (4.18)$$

En passant à la limite pour $j \rightarrow +\infty$, on trouve :

- $\|\hat{x} - x^{k_j}\| \rightarrow \hat{\beta}$
car, si une suite converge, toute sous-suite converge vers la même limite;
- $\|\bar{x} - x^{k_j}\| \rightarrow \bar{\beta}$
car, si une suite converge, toute sous-suite converge vers la même limite;

$$\bullet \quad x^{k_j} \rightharpoonup \bar{x} \quad \Leftrightarrow \quad \forall z \in X : \quad \langle x^{k_j}, z \rangle \rightarrow \langle \bar{x}, z \rangle$$

$$\text{d'où } \langle \hat{x} - \bar{x}, \bar{x} - x^{k_j} \rangle \rightarrow \langle \hat{x} - \bar{x}, \bar{x} - \bar{x} \rangle = 0 ;$$

ainsi, (4.18) donne

$$\hat{\beta}^2 = \|\bar{x} - \hat{x}\|^2 + \bar{\beta}^2 . \quad (4.19)$$

En sommant (4.17) et (4.19), on obtient

$$\begin{aligned} \bar{\beta}^2 + \hat{\beta}^2 &= 2 \|\bar{x} - \hat{x}\|^2 + \hat{\beta}^2 + \bar{\beta}^2 \Leftrightarrow 2 \|\bar{x} - \hat{x}\|^2 = 0 \\ &\Leftrightarrow \|\bar{x} - \hat{x}\| = 0 \\ &\Leftrightarrow \hat{x} = \bar{x} . \end{aligned}$$

\bar{x} est donc bien l'unique point d'accumulation faible de $\{x^k\}$.

On a ainsi montré que, si $\mathcal{S} \neq \emptyset$, alors la suite $\{x^k\}$ converge faiblement vers une solution du problème.

2ème cas : $\mathcal{S} = \emptyset$

Supposons par l'absurde que la suite $\{x^k\}$ est bornée.

Par le Pas # 2 de l'algorithme,

$$x^{k+1} = x^k - c_k v^k \quad \Leftrightarrow \quad v^k = \frac{1}{c_k} (x^k - x^{k+1})$$

où, par hypothèse,

$$c_k \geq \bar{c} > 0 , \quad \forall k .$$

Ainsi, les suites $\{c_k v^k\}$ et $\{v^k\}$ sont également bornées.

De plus,

$$\begin{aligned} \|c_k v^k\| &= \|-c_k v^k\| \\ &= \|y^k - x^k - (c_k v^k + y^k - x^k)\| \\ &\geq \left| \|y^k - x^k\| - \|c_k v^k + y^k - x^k\| \right| \end{aligned}$$

$$\begin{aligned}
&\geq \left| \|y^k - x^k\| - \sigma \|y^k - x^k\| \right| \quad [\text{cf. Pas \# 1 de l'algorithme}] \\
&\quad \|c_k v^k + y^k - x^k\|^2 + 2 c_k \varepsilon_k \leq \sigma^2 \|y^k - x^k\|^2 \\
&\quad \text{comme tous les termes sont positifs ou nuls} \\
&\quad \text{on en déduit immédiatement que} \\
&\quad \|c_k v^k + y^k - x^k\| \leq \sigma \|y^k - x^k\| \\
&= (1 - \sigma) \|y^k - x^k\|. \quad [\text{cf. } \sigma \in [0, 1[\Rightarrow 0 < 1 - \sigma \leq 1]
\end{aligned}$$

Par conséquent, $\{y^k\}$ est aussi bornée.

Comme T est maximal monotone, le domaine $D(T)$ n'est pas vide.

Soient alors

- $\bar{z} \in D(T) \neq \emptyset$;
- $R > \max\{\|\bar{z}\|, \sup_k \|x^k\|, \sup_k \|y^k\|\}$
 (R est bien un nombre fini car les suites $\{x^k\}$ et $\{y^k\}$ sont bornées, donc le sup existe);
- $B \stackrel{\text{déf}}{=} \{z \in X \mid \|z\| \leq 2R\}$.

Considérons le cône normal de l'ensemble convexe B , c'est-à-dire l'opérateur N_B défini par

$$N_B(x) = \{\nu \mid \langle \nu, b - x \rangle \leq 0, \forall b \in B\}.$$

On a

$$N_B = \partial I_B$$

où I_B est la fonction indicatrice de l'ensemble B définie par

$$\begin{aligned}
I_B : X &\rightarrow \mathbb{R} \cup \{+\infty\}, \\
x &\mapsto I_B(x) = \begin{cases} 0 & \text{si } x \in B, \\ +\infty & \text{si } x \notin B. \end{cases}
\end{aligned}$$

En effet,

$$x^* \in \partial I_B(x) \Leftrightarrow I_B(z) \geq I_B(x) + \langle x^*, z - x \rangle, \quad \forall z \in X. \quad [\text{cf. Définition 1.8}]$$

Cette inégalité est impossible si $x \notin B$.

Il faut donc que $I_B(x) = 0$.

De plus, elle doit seulement être vérifiée pour tout $z \in B$ car elle sera toujours vraie si $z \notin B$.

Ainsi,

$$\begin{aligned} x^* \in \partial I_B(x) &\Leftrightarrow 0 \geq 0 + \langle x^*, z - x \rangle, \quad \forall z \in B \\ &\Leftrightarrow x^* \in N_B(x). \end{aligned} \quad [\text{par définition}]$$

Or, on peut montrer que la fonction indicatrice I_B est convexe si la partie B est convexe; elle est propre si B n'est pas vide et elle est semi-continue inférieurement si B est une partie fermée.

Comme $B = \{z \in H \mid \|z\| \leq 2R\}$ satisfait ces hypothèses, $N_B = \partial I_B$ est donc le sous-différentiel d'une fonction convexe propre, semi-continue inférieurement et est donc un opérateur maximal monotone en vertu de la Proposition 1.2.

Comme

$$\begin{aligned} \bar{z} &\in (D(T) \cap \overset{\circ}{D}(N_B)) \\ &[\text{cf. } \|\bar{z}\| < R \text{ par définition de } R, \text{ d'où } \bar{z} \in \overset{\circ}{B}] \end{aligned}$$

on a

$$(D(T) \cap \overset{\circ}{D}(N_B)) \neq \emptyset$$

et, par conséquent, la somme des deux opérateurs maximaux monotones reste maximale monotone [on peut trouver ce résultat de Rockafellar dans [12]],

c'est-à-dire $T' = T + N_B$ est maximal monotone.

De plus, le domaine de T' est borné

et, par conséquent, il admet un zéro [on peut trouver ce résultat de Rockafellar dans [13], Proposition 2],

c'est-à-dire le problème :

$$\text{trouver } x \in X \text{ tel que } 0 \in T'(x)$$

admet une solution.

Montrons maintenant que

$$\forall k : \quad v^k \in (T')^{\varepsilon_k}(y^k).$$

Soient $z \in X$ et $w' \in T'(z)$ quelconques.

Alors, par définition de $T' = T + N_B$, il existe $w, v \in X$ tels que

$$w' = w + v \quad \text{avec } w \in T(z) \text{ et } v \in N_B(z).$$

Or,

$$\begin{aligned} y^k \in B, \quad \forall k. \quad & [\text{car } \sup_k \|y^k\| < R \text{ par définition de } R, \\ & \text{d'où } \|y^k\| < R, \forall k \text{ et } \|y^k\| < 2R, \forall k] \end{aligned}$$

Ainsi, par définition du cône normal :

$$v \in N_B(z) \Leftrightarrow \langle v, b - z \rangle \leq 0, \quad \forall b \in B,$$

on a

$$\begin{aligned} \langle v, y^k - z \rangle &\leq 0, \quad \forall k, \\ \Leftrightarrow \langle v, z - y^k \rangle &\geq 0, \quad \forall k. \end{aligned} \tag{4.20}$$

Comme, par le Pas # 1 de l'algorithme, $v^k \in T^{\varepsilon_k}(y^k)$, on a, par la Définition 1.10,

$$\langle u - v^k, x - y^k \rangle \geq -\varepsilon_k, \quad \forall x \in X, \forall u \in T(x),$$

donc, en particulier pour $x = z$ et $u = w \in T(z)$, on a

$$\langle w - v^k, z - y^k \rangle \geq -\varepsilon_k. \tag{4.21}$$

D'où

$$\begin{aligned} \langle w' - v^k, z - y^k \rangle &= \langle w - v^k, z - y^k \rangle + \langle v, z - y^k \rangle && [\text{cf. } w' = w + v] \\ &\geq -\varepsilon_k + \langle v, z - y^k \rangle && [\text{cf. (4.21)}] \\ &\geq -\varepsilon_k, \quad \forall k; && [\text{cf. (4.20)}] \end{aligned}$$

comme $z \in X$ et $w' \in T'(z)$ étaient choisis quelconques, on a, par la Définition 1.10,

$$\forall k : \quad v^k \in (T')^{\varepsilon_k}(y^k).$$

Ainsi, les suites $\{x^k\}$, $\{(y^k, v^k)\}$ et $\{\varepsilon_k\}$ peuvent être considérées comme les suites générées par l'algorithme appliqué au problème :

$$\text{trouver } x \text{ tel que } 0 \in T'(x)$$

dont on vient de voir qu'il admet des solutions.

Mais alors, en vertu de la première partie de ce théorème, la suite $\{x^k\}$ converge faiblement vers une solution x' de ce problème et on a :

$$0 \in T'(x') = T(x') + N_B(x') . \quad (4.22)$$

Or, par définition de R ,

$$\sup_k \|x^k\| < R ,$$

d'où

$$\|x^k\| < R , \quad \forall k ,$$

et

$$\|x'\| \leq R , \text{ [en passant à la limite et par continuité de la norme]}$$

de sorte que $x' \in \overset{\circ}{B}$, ce qui implique que

$$N_B(x') = \{\nu \mid \langle \nu, b - x' \rangle \leq 0 , \forall b \in B\} = \{0\}$$

et on aurait

$$0 \in T(x') , \quad [\text{cf. (4.22)}]$$

c'est-à-dire $x' \in \mathcal{S}$,

ce qui contredit l'hypothèse que $\mathcal{S} = \emptyset$.

On a donc bien que, si $\mathcal{S} = \emptyset$, alors $\{x^k\}$ n'est pas bornée. ■

Théorème 4.2

*Si T^{-1} est Lipschitz continu autour de zéro,
alors la suite $\{x^k\}$ générée par l'algorithme hybride extragradient-point proximal ap-
proché converge linéairement vers la solution du problème*

$$\text{trouver } x \text{ tel que } 0 \in T(x) .$$

Preuve : T^{-1} est Lipschitz continu autour de zéro si et seulement si

$\exists \delta > 0$, $\exists L > 0$ tels que

$$\forall v : \|v - 0\| \leq \delta , \quad y \in T^{-1}(v) \Rightarrow \|y (\in T^{-1}(v)) - x^* (\in T^{-1}(0))\| \leq L \|v - 0\| ,$$

i.e.

$$\forall v : \|v\| \leq \delta , \quad v \in T(y) \Rightarrow \|y - x^*\| \leq L \|v\| .$$

Ceci montre que, sous ces hypothèses, on peut parler de la solution (unique) du problème, car si $0 \in T(y)$, on a évidemment $\|0\| \leq \delta$ et donc, par la Lipschitz continuité,

$$0 \leq \|y - x^*\| \leq L \|0\| = 0 ,$$

ce qui implique que

$$y \equiv x^* .$$

Pour tout k , notons z^k la solution exacte du sous-problème prox (qui existe toujours en vertu du Théorème 1.1 car T est maximal monotone), c'est-à-dire

$$\begin{aligned} 0 &\in c_k T(z^k) + (z^k - x^k) \\ \Leftrightarrow \quad & - \frac{z^k - x^k}{c_k} \in T(z^k) . \end{aligned} \tag{4.23}$$

en vertu du Lemme 4.1, on a

$$\|z^k - y^k\| \leq \sigma \|y^k - x^k\| . \tag{4.24}$$

De plus,

$$\begin{aligned} \|z^k - x^k\| &= \|z^k - y^k + y^k - x^k\| \\ &\leq \|z^k - y^k\| + \|y^k - x^k\| \quad [\text{inégalité triangulaire}] \\ &\leq (1 + \sigma) \|y^k - x^k\| . \end{aligned} \tag{4.25}$$

[cf. (4.24)]

Ainsi,

$$\begin{aligned} \left\| - \frac{z^k - x^k}{c_k} \right\| &= \frac{1}{c_k} \|z^k - x^k\| \\ &\leq \frac{1 + \sigma}{c_k} \|y^k - x^k\| \quad [\text{cf. (4.25)}] \\ &\leq \frac{1 + \sigma}{\bar{c}} \|y^k - x^k\| . \quad [\text{cf. } c_k \geq \bar{c} > 0 , \forall k] \end{aligned}$$

On sait, par le Corollaire 4.1, 2.(b), que

$$\|y^k - x^k\| \xrightarrow{k \rightarrow +\infty} 0 ,$$

donc, par l'inégalité précédente,

$$\left\| -\frac{z^k - x^k}{c_k} \right\| \xrightarrow{k \rightarrow +\infty} 0$$

et, par conséquent,

$$\exists k_0 \text{ tel que } \forall k \geq k_0 : \left\| -\frac{z^k - x^k}{c_k} \right\| \leq \delta .$$

Comme, de plus,

$$-\frac{z^k - x^k}{c_k} \in T(z^k) , \quad [\text{cf. (4.23)}]$$

c'est-à-dire

$$z^k \in T^{-1} \left(-\frac{z^k - x^k}{c_k} \right) , \quad [\text{cf. Définition 1.4}]$$

on déduit de la Lipschitz continuité de T^{-1} autour de zéro que, pour $k \geq k_0$,

$$\|z^k - x^k\| \leq L \left\| -\frac{z^k - x^k}{c_k} \right\| = \frac{L}{c_k} \|z^k - x^k\| . \quad (4.26)$$

De plus,

$$\begin{aligned} \|x^* - x^k\| &= \|x^* - z^k + z^k - x^k\| \\ &\leq \|x^* - z^k\| + \|z^k - x^k\| && [\text{inégalité triangulaire}] \\ &\leq \frac{L}{c_k} \|z^k - x^k\| + \|z^k - x^k\| && [\text{cf. (4.26)}] \\ &\leq \left(1 + \frac{L}{c}\right) \|z^k - x^k\| && [\text{cf. } c_k \geq \bar{c} > 0, \forall k] \\ &\leq \left(1 + \frac{L}{c}\right) (1 + \sigma) \|y^k - x^k\| . && [\text{cf. (4.25)}] \end{aligned}$$

En combinant cette majoration avec le Lemme 4.2, on obtient, pour $k \geq k_0$:

$$\|x^* - x^k\|^2 - \|x^* - x^{k+1}\|^2 \geq (1 - \sigma^2) \|y^k - x^k\|^2 ,$$

$$\begin{aligned}
\|x^* - x^{k+1}\|^2 &\leq \|x^* - x^k\|^2 - (1 - \sigma^2) \left[\frac{1}{\left(1 + \frac{L}{\bar{c}}\right)^2 (1 + \sigma)} \|x^* - x^k\| \right]^2 \\
&= \left[1 - \frac{(1 - \sigma)(1 + \sigma)}{\left(1 + \frac{L}{\bar{c}}\right)^2 (1 + \sigma)^2} \right] \|x^* - x^k\|^2 \\
&= \left[1 - \frac{1 - \sigma}{\left(1 + \frac{L}{\bar{c}}\right)^2 (1 + \sigma)} \right] \|x^* - x^k\|^2.
\end{aligned}$$

Comme $\sigma \in [0, 1[$, $L > 0$ et $\bar{c} > 0$, on a toujours

$$0 < \frac{1 - \sigma}{\left(1 + \frac{L}{\bar{c}}\right)^2 (1 + \sigma)}$$

et

$$\frac{1 - \sigma}{\left(1 + \frac{L}{\bar{c}}\right)^2 (1 + \sigma)} \leq 1;$$

en effet,

$$(1 - \sigma) \leq \left(1 + \frac{L}{\bar{c}}\right)^2 (1 + \sigma)$$

car

$$\underbrace{-2}_{\leq 0} \sigma \leq \underbrace{\left(2 \frac{L}{\bar{c}} + \frac{L^2}{\bar{c}^2}\right)}_{\geq 0} \underbrace{(1 + \sigma)}_{\geq 0}.$$

En posant

$$\alpha^2 \equiv \left[1 - \frac{1 - \sigma}{\left(1 + \frac{L}{\bar{c}}\right)^2 (1 + \sigma)} \right],$$

on obtient donc : $\exists 0 \leq \alpha < 1$ tel que $\forall k \geq k_0$,

$$\|x^* - x^{k+1}\| \leq \alpha \|x^* - x^k\|.$$

D'où $\{x^k\}$ converge linéairement vers x^* .

Remarquons encore que, si on laisse tendre la tolérance d'erreur σ_k vers 0 et le paramètre de régularisation c_k vers l'infini, on a :

$$\alpha_k^2 = \left[1 - \frac{\overbrace{(1 - \sigma_k)}^{\rightarrow 1}}{\underbrace{\left(1 + \frac{L}{\underbrace{c_k}_{\rightarrow 0}}\right)^2}_{\rightarrow 1} \underbrace{(1 + \sigma_k)}_{\rightarrow 1}} \right] \rightarrow 0.$$

on obtient donc : $\exists \alpha_k \xrightarrow{>} 0$ tel que, pour tout $k \geq k_0$,

$$\|x^* - x^{k+1}\| \leq \alpha_k \|x^* - x^k\| ,$$

c'est-à-dire la convergence devient superlinéaire. ■

4.4 Application : méthode de Newton généralisée

Dans cette section, on va décrire comment on peut utiliser la méthode hybride extragradient-point proximal approchée pour construire une méthode de type Newton globalement convergente.

Soient $X = \mathbb{R}^n$ et $T = F$

avec $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ une fonction monotone et continûment différentiable, dont la jacobienne est Lipschitz continue, c'est-à-dire

$$\exists \gamma > 0 \quad \text{tel que} \quad \|\nabla F(x) - \nabla F(y)\| \leq \gamma \|x - y\| , \quad \forall x, y \in \mathbb{R}^n .$$

Notre problème consiste donc à résoudre le système d'équations

$$F(x) = 0 .$$

La clé de la méthode qu'on va décrire est qu'un seul pas Newton appliqué au sous-problème prox

$$0 = c_k F(x) + (x - x^k)$$

suffit pour assurer le critère de tolérance d'erreur à la k -ième itération de l'algorithme hybride, pourvu que l'on choisisse

$$c_k \leq \frac{1}{\sqrt{\gamma \|F(x^k)\|}} \quad \text{et} \quad \sigma \in \left[\frac{1}{2}, 1\right[;$$

autrement dit, si

$$s^k = -(c_k \nabla F(x^k) + I)^{-1} c_k F(x^k) \quad [\text{direction Newton}]$$

et

$$y^k = x^k + s^k ,$$

alors

$$\|c_k F(y^k) + y^k - x^k\| \leq \sigma \|y^k - x^k\| .$$

[critère de tolérance d'erreur de l'algorithme hybride 4.2.4

avec $\varepsilon_k = 0$ et $T = F$ univoque]

En effet,

$$\begin{aligned} & \|F(y^k) - F(x^k) - \nabla F(x^k) (y^k - x^k)\| \\ &= \|[F(x^k + t s^k)]_0^1 - \nabla F(x^k) s^k\| \quad [\text{cf. } y^k = x^k + s^k] \\ &= \left\| \int_0^1 \nabla F(x^k + t s^k) s^k dt - \nabla F(x^k) s^k \right\| \\ &= \left\| \int_0^1 [\nabla F(x^k + t s^k) s^k - \nabla F(x^k) s^k] dt \right\| \\ &\leq \int_0^1 \|\nabla F(x^k + t s^k) s^k - \nabla F(x^k) s^k\| dt \\ &\leq \int_0^1 \|\nabla F(x^k + t s^k) - \nabla F(x^k)\| \|s^k\| dt \quad [\text{Cauchy-Schwarz}] \\ &\leq \int_0^1 \gamma \|t s^k\| \|s^k\| dt \quad [\nabla F \text{ est Lipschitz continue}] \\ &= \gamma \|s^k\|^2 \int_0^1 t dt \\ &= \gamma \|s^k\|^2 \left[\frac{1}{2} t^2 \right]_0^1 \\ &= \frac{1}{2} \gamma \|s^k\|^2 . \end{aligned}$$

Or,

$$\begin{aligned} & -c_k F(x^k) - c_k \nabla F(x^k) s^k \\ &= -c_k F(x^k) + c_k \nabla F(x^k) (c_k \nabla F(x^k) + I)^{-1} c_k F(x^k) \quad [\text{par définition de } s^k] \\ &= -c_k F(x^k) + (c_k \nabla F(x^k) + I) (c_k \nabla F(x^k) + I)^{-1} c_k F(x^k) \\ &\quad - (c_k \nabla F(x^k) + I)^{-1} c_k F(x^k) \\ &= -c_k F(x^k) + c_k F(x^k) - (c_k \nabla F(x^k) + I)^{-1} c_k F(x^k) \\ &= s^k . \end{aligned} \tag{4.27}$$

Ainsi,

$$\begin{aligned} \|F(y^k) - F(x^k) - \nabla F(x^k) (y^k - x^k)\| &\leq \frac{1}{2} \gamma \|s^k\|^2 \\ \frac{1}{c_k} \|c_k F(y^k) - c_k \underbrace{F(x^k) - c_k \nabla F(x^k) s^k}_{\geq 0}\| &\leq \frac{1}{2} \gamma \|s^k\|^2 \\ \frac{1}{c_k} \|c_k F(y^k) + s^k\| &\leq \frac{1}{2} \gamma \|s^k\|^2 . \quad [\text{cf. (4.27)}] \end{aligned}$$

Comme F est monotone et continûment différentiable, ∇F est semi-défini positif et (4.27) donne

$$\langle s^k, s^k \rangle = -c_k \langle F(x^k), s^k \rangle - \underbrace{c_k}_{>0} \underbrace{\langle \nabla F(x^k) s^k, s^k \rangle}_{\geq 0} ,$$

D'où

$$\begin{aligned} \|s^k\|^2 &\leq -c_k \langle F(x^k), s^k \rangle \\ &\leq c_k \|F(x^k)\| \|s^k\| . \quad [\text{Cauchy-Schwarz}] \end{aligned}$$

Ainsi, on obtient successivement

$$\begin{aligned} \frac{1}{c_k} \|c_k F(y^k) + s^k\| &\leq \frac{1}{2} \gamma c_k \|F(x^k)\| \|s^k\| , \\ \|c_k F(y^k) + y^k - x^k\| &\leq \frac{1}{2} \gamma c_k^2 \|F(x^k)\| \|y^k - x^k\| \quad [\text{cf. } y^k = x^k + s^k] \\ &\leq \frac{1}{2} \gamma \frac{1}{\gamma \|F(x^k)\|} \|F(x^k)\| \|y^k - x^k\| \quad [\text{cf. choix de } c_k] \\ &= \frac{1}{2} \|y^k - x^k\| \\ &\leq \sigma \|y^k - x^k\| , \quad [\text{cf. choix de } \sigma] \end{aligned}$$

ce qu'il fallait montrer.

Notre méthode de Newton régularisée dont chaque itération consiste en un pas Newton suivi d'un pas extragradient s'écrit donc :

Initialisation :

Choisir $x^0 \in \mathbb{R}^n$,

$\sigma \in \left[\frac{1}{2}, 1\right[$,

$\bar{c} > 0$.

Itération : pour $k = 0, 1, \dots$:

Pas # 1

Choisir $\bar{c} \leq c_k \leq \frac{1}{\sqrt{\gamma \|F(x^k)\|}}$

et calculer $s^k = -(c_k \nabla F(x^k) + I)^{-1} c_k F(x^k)$,

$$y^k = x^k + s^k.$$

Pas # 2

Définir $x^{k+1} = x^k - c_k F(y^k)$.

En vertu du Théorème 4.1, on a alors immédiatement que **toute** la suite générée par cette méthode converge **globalement** (ici, on a même la convergence forte car on travaille dans \mathbb{R}^n) vers une solution du système d'équations, pourvu qu'une solution existe, sans imposer des hypothèses de régularité supplémentaires. Ce résultat paraît d'autant plus important qu'on sait que d'autres méthodes de Newton globalisées, basées sur des fonctions de mérite par exemple, doivent faire l'hypothèse que les ensembles de niveau (et donc l'ensemble des solutions) sont bornés pour prouver la convergence de **sous-suites** vers des points stationnaires de ces fonctions qui ne sont pas nécessairement des solutions de $F(x) = 0$. Pour garantir la convergence **locale** de toute la suite, ces méthodes doivent en plus imposer l'unicité de la solution.

Notons encore que, si γ n'est pas connu, on peut chercher un c_k adéquat par une recherche linéaire de type Armijo. Si on suppose que ∇F n'est pas singulier à la solution, on peut même prendre pour k suffisamment grand

$$c_k = \frac{1}{\sqrt{\|F(x^k)\|}}$$

et aucune recherche linéaire n'est nécessaire. On peut en plus montrer que sous cette condition, la convergence est superlinéaire.

(On peut trouver les détails techniques de ces résultats dans [16].)

Bibliographie

- [1] BREZIS Haim : *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*. North Holland, 1973.
- ✓ [2] BURACHIK R.S., IUSEM A.N. & SVAITER B.F. : Enlargement of monotone operators with applications to variational inequalities. *Set-Valued Analysis* **5**, pp. 159–180, 1997.
- ✓ [3] BURACHIK R.S., SAGASTIZÁBAL C.A. & SVAITER B.F. : ε -Enlargements of maximal monotone operators: Theory and applications. In: FUKUSHIMA M. & L. QI Editors: *Reformulation-Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, pp. 25–44. Kluwer Academic Publishers, 1998.
- ✓ [4] BURACHIK R.S., SAGASTIZÁBAL C.A. & SVAITER B.F. : Bundle methods for maximal monotone operators, submitted, 1997.
- [5] CLARKE Frank H. : Optimization and nonsmooth analysis. *Canadian Mathematical Society Series of Monographs and Advanced Texts*. A Wiley Interscience Publication.
- [6] ECKSTEIN Jonathan & BERTSEKAS Dimitri, P. : On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming* **55**, pp. 293–318, 1992.
- [7] HIRIART-URRUTY Jean-Baptiste & LEMARÉCHAL Claude : Convex analysis and minimization algorithms. *Grund. der Math. Wissenschaften*. Springer Verlag, 1993 (2 volumes).
- [8] IUSEM A.N., SVAITER B.F. & TEBOULE M. : Entropy-like methods in convex programming. *Mathematics of Operations Research* **19**, pp. 790–814, 1994.
- [9] KORPELEVICH G.M. : The extragradient method for finding saddle points and other problems. *Matecon* **12**, pp. 747–756, 1976.
- [10] ROCKAFELLAR R.T. : Local boundedness of nonlinear monotone operators. *Michigan Math. Journal* **16**, pp. 397–407, 1969.

- [11] ROCKAFELLAR R.T.: *Convex Analysis*. Princeton University Press, 1970.
- [12] ROCKAFELLAR R.T.: On the maximality of sums of nonlinear monotone operators. *Transactions of the American Mathematical Society* **149**, pp. 75–88, 1970.
- [13] ROCKAFELLAR R.T.: Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization* **14**, pp. 877–898, 1976.
- ✗ [14] SOLODOV M.V. & SVAITER B.F.: A hybrid projection-proximal point algorithm. *Journal of Convex Analysis*, to appear.
- ✗ [15] SOLODOV M.V. & SVAITER B.F.: A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Analysis*, submitted, 1998.
- ✗ [16] SOLODOV M.V. & SVAITER B.F.: A globally convergent inexact Newton method for systems of monotone equations. In: FUKUSHIMA M. AND L. QI Editors: *Reformulation-Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, pp. 355–369. Kluwer Academic Publishers, 1998.
- [17] STRODIOT J.-J. & NGUYEN V.H.: On the numerical treatment of the inclusion $0 \in \partial f(x)$. In: MOREAU J.J., PANAGIOTOPOULOS & STRAG G.: *Topics in Nonsmooth Mechanics*, Chap. VII, pp. 267–294. Birkhäuser Verlag, 1988.

Table des matières

Introduction	1
Chapitre 1 :	
Opérateurs maximaux monotones	3
1.1 Définitions et propriétés	3
1.2 ε -élargissement d'un opérateur maximal monotone	12
1.3 Propriétés de T^ε	14
Chapitre 2 :	
Algorithme conceptuel	36
2.1 Introduction et motivation	36
2.2 L'algorithme conceptuel	38
2.3 Analyse de convergence	39
Chapitre 3 :	
Algorithme implémentable	52
3.1 Introduction et motivation	52
3.2 L'algorithme implémentable	60
3.3 Analyse de convergence	64

Chapitre 4 :

Méthode hybride extragradient-point proximal approchée	80
4.1 Introduction et motivation	80
4.1.1 Le reste	81
4.1.2 L' ε -élargissement	82
4.1.3 Première approche	83
4.1.4 Deuxième approche	84
4.2 La méthode hybride extragradient-point proximal approchée	85
4.2.1 Comparaison avec 4.1.2	86
4.2.2 Comparaison avec 4.1.3	90
4.2.3 Propriété des solutions approchées	92
4.2.4 L'algorithme hybride extragradient-point proximal approché	94
4.2.5 Remarques	95
4.3 Analyse de convergence	100
4.4 Application : méthode de Newton généralisée	119

Bibliographie